Springfield-Chicopee School Districts Striving Readers (SR) Program

Year 2 Report: Evaluation of Implementation and Impact

April 2009

Prepared by:
Research & Evaluation Division
The Education Alliance at Brown University

222 Richmond Street, Suite 300
Providence, Rhode Island 02903-4226

Phone: 401.274.9548
Fax: 401.421.7650
E-mail: information@alliance.brown.edu
Web: www.alliance.brown.edu

Springfield-Chicopee School Districts Striving Readers (SR) Program

Year 2 Report: Evaluation of Implementation and Impact

April 2009

Prepared for:
Office of Elementary and Secondary Education, U.S. Department of Education

Prepared by:
Research & Evaluation Division
The Education Alliance at Brown University

THE EDUCATION ALLIANCE at Brown University

EQUITY AND EXCELLENCE FOR ALL SCHOOLS

Since 1975, The Education Alliance, a department at Brown University, has helped the education community improve America's schools. We provide applied research, technical assistance, and informational resources to connect research and practice, build knowledge and skills, and meet critical needs in the field.

With offices located in Providence, Rhode Island, adjacent to the Brown University campus, and a dedicated team of more than 50 skilled professionals, collaborators, and partners, we provide services and resources to K-12 schools and districts across the country and beyond. As we work with educators, we customize our programs to the specific needs of our clients.

Our Web site (www.alliance.brown.edu) describes our work and provides extensive information and resources about education reform. Information about all Alliance programs and services is available by contacting:

The Education Alliance
at Brown University
222 Richmond Street, Suite 300
Providence, RI 02903-4226

Phone: 800.521.9550
Fax: 401.421.7650
E-mail: information@alliance.brown.edu
Web: www.alliance.brown.edu

Report Authors: Kimberley Sprague, Ivana Zuliani, Colleen Zaller, Anita Kite

---

RESEARCH AND EVALUATION DIVISION

The Education Alliance's Research and Evaluation division engages a wide range of quantitative and qualitative strategies to provide educators with information they can use to improve programs, inform school reform activities, and facilitate change. Methods include experimental designs using randomized control trials to examine causal relationships and quasi-experimental designs employing well-specified program and comparison groups to understand program effects. Our approach considers multiple perspectives and local contexts to ensure that educators can apply our recommendations to their specific problems. We have conducted assessments of school change models, online professional development offerings, school support organizations, and ESL and bilingual programs. We also evaluate how assessment strategies inform professional development and classroom practice in terms of national standards. For further information about the services of this division, contact Deborah Collins, Director of Research and Evaluation, at 800.521.9550, Ext 260, or email reseval@alliance.brown.edu.

ACKNOWLEDGEMENTS

The authors recognize the "above-and-beyond" contributions and commitment of our partners in the Springfield and Chicopee Public School Districts (administrators, teachers, and program staff). A special thanks, in particular, to the phenomenal Striving Readers Implementation Team who worked tirelessly to ensure this study would contribute to the research base in the field of education. This study has benefited from the commitment and energy of the Striving Readers Team including Ann Ferriter, Matt Rigney, Sheila Hoffman, and Justin Hurst who facilitated the implementation of the three interventions specified in the project with the research always in mind; facilitated access to classrooms and teachers as well as school and district staff for interviews; and responded willingly to requests for data documenting their work.

The authors acknowledge the significant contributions of the Project Officer Marcia Kingman at the Office of Elementary and Secondary Education, Stefanie Schmidt at the Institute for Education Sciences, Barbara Goodson, Ryoko Yamaguchi, Cris Price, and Beth Boulay at Abt Associates, Inc. (all Abt technical assistance team members), and Julie Meltzer at the Center for Resource Management-Public Consulting Group, Inc. In addition, formative and substantial (measurable) contributions were made by Jennifer Borman and Bob St. Pierre for which the authors are grateful.

Thanks to our Research and Evaluation Director Deborah Collins for her support and other current and former colleagues as well: Chandra Haislet, Laurie Phillips, Sarah Cussler, Joan Ford, Hardeek Shah, Kim Burke, Julia Frizzell, Martin Huntley, Sunitha Appikatla, Stephanie Feger, Audrey Chait, Steven Lacey, Roy Echeverria and Erika Taylor. Any omission of acknowledgement is solely the responsibility of the authors.

TABLE OF CONTENTS

LIST OF EXHIBITS

Executive Summary

This second report for the Striving Readers program presents implementation and impact findings for the first two years of the grant implemented by the Springfield and Chicopee Public Schools. The U.S. Department of Education (ED), Office of Elementary and Secondary Education (OESE), has funded the implementation of the Striving Readers program and provided oversight. The Institute of Education Sciences (IES) at ED has provided oversight for the evaluation component and the ED-contracted Striving Readers technical assistance provider is Abt Associates, Inc.[1] The Striving Readers grant required the implementation of both targeted and whole school literacy interventions. In addition, this grant required the inclusion of rigorous evaluation components and implementation studies.

In the Springfield and Chicopee Public School Districts, five high schools (three in Springfield and two in Chicopee) are implementing two targeted interventions—both developed using scientifically-based research (SBR) to promote the reading skills of struggling readers—as well as a whole school intervention developed to promote content-area literacy skills throughout the student population. The targeted interventions are: (1) READ 180 Enterprise Edition (Scholastic, Inc.) and (2) Strategic Instruction Model (SIM) Xtreme Reading (University of Kansas, Center for Research on Learning). Both targeted interventions were provided as a supplement to the regular English language arts curriculum in participating schools. The school-wide intervention is the Strategic Instruction Model, Content Enhancement Routines for Teachers (SIM-CERT), which along with Xtreme Reading is a part of the University of Kansas' Content Literacy Continuum.

[1] The authors acknowledge the significant contributions of the Project Officer Marcia Kingman at the Office of Elementary and Secondary Education, Stefanie Schmidt at the Institute for Education Sciences, Barbara Goodson, Ryoko Yamaguchi, Cris Price, and Beth Boulay at Abt Associates, Inc. (all Abt technical assistance team members), and Julie Meltzer at the Center for Resource Management-Public Consulting Group, Inc. In addition, formative and substantial (measurable) contributions were made by Jennifer Borman and Bob St. Pierre for which the authors are grateful. Finally, the most critical recognition is reserved for our partners in the Springfield and Chicopee Public School Districts and special thanks, in particular, to the phenomenal Striving Readers Implementation Team who worked tirelessly to ensure this study would contribute to the field.

Targeted Evaluation Design Overview

The targeted impact study was rigorously designed to determine whether or not targeted intervention participation improves the reading achievement of struggling (striving) readers. A randomized controlled trial (RCT) was employed to provide estimates of the "true" effect of the interventions on reading achievement. Eligible 9th grade students[2] were randomly assigned to participate in one of the two supplemental programs (READ 180 or Xtreme Reading) or to "business-as-usual," defined as any support normally provided in the district to students struggling in reading.[3] Eligible teachers were also randomly assigned to teach students who were randomly assigned to READ 180, Xtreme Reading, or the Control group. The unit of randomization for the analysis was the student so power for analysis was calculated at this level. Average reading achievement scores[4] of students in each of the two interventions were compared to the scores of students in control classrooms, pooled across sites and study years.

The targeted implementation study was designed to describe the context in which the interventions and any concomitant effects were, observed. The evaluation focused on the extent to which the interventions were implemented "on-model" and also sought to describe the general implementation context such that it could be used to inform the interpretation of outcomes. For this study, "on-model" is the extent to which the targeted intervention is implemented according to the developers' and districts' specifications and plans.[5] Implementation levels were established to characterize implementation context and its complexity in a clear and understandable way. As a result, the levels provide a gauge by which to judge implementation context in relationship to any observed effects.

[2] As per district plans, the effectiveness of each targeted intervention is tested in 9th grade. The *Scholastic Reading Inventory* (SRI) was the district screening tool (chosen because systems were already in place to implement it).
[3] In the absence of supplemental support, students participate in other electives.
[4] The primary outcome data on student reading achievement were collected by the districts using the *Stanford Diagnostic Reading Test, 4th edition* (SDRT-4).
[5] Project Officer communication, November 15, 2006.

Targeted Implementation Study

Scores assessing fidelity to model implementation were assigned for each intervention. The five components of the fidelity scores were based on inputs, outputs, and indirect elements: (1) professional development (input); (2) materials, technology, assessments (input); (3) classroom organization, structure, context (input); (4) classroom model including instruction practices (output); and (5) student behavior and classroom management (indirect). Implementation levels were defined as follows: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%). Differences in implementation levels between Year 1 and Year 2, particularly for the classroom model, may in part be due to the refined specificity of measures related to the interventions (based on previously unavailable information). The refined indicators made the scoring more stringent during this second round of analysis. As stated previously, such measures were added to more accurately capture model fidelity. Therefore, scores should be interpreted with caution because they are influenced by weighted subcomponent scores but also because they were based on classroom observations which represent "snap-shots" in time.

READ 180 Implementation. Professional development ratings were either moderate or adequate for all five teachers. When Year 2 ratings were compared to that of Year 1, they were more consistent and positive. Most teachers implemented the intervention at a moderate or adequate level in terms of classroom model: two of the five teachers were rated adequate, the highest level of implementation, and two were rated moderate (defined as implementing a majority of model components, a majority of the time), and the remaining teacher was rated low, indicating that the appropriate level of implementation for the classroom model was not achieved (note that this teacher was replaced in Year 3). Two of the five (40%) teachers received ratings of adequate at the end of Year 2 in the implementation of *both* inputs and classroom model (two of six in Year 1 or 33% received this rating). The remaining three out of the five teachers were rated at mixed levels for both inputs and classroom model.

Xtreme Implementation. Two of the three model inputs (professional development and class structure) were rated as adequate in Year 2 for four teachers, which influenced the overall input ratings. More specifically, two teachers achieved a rating of adequate in Year 2 while four teachers achieved an adequate rating in Year 1.[6] Three out of the five teachers received a rating of adequate at the end of Year 1 in the implementation of *both* inputs and classroom model. In Year 2, none of the teachers received "adequate" ratings for both of these categories. While two teachers received moderate ratings, indicating a majority of component indicators were observed, the remaining three received mixed ratings for both inputs and the classroom model.

Targeted Impact Study

Business as Usual. The two components of "business as usual" for striving readers include: (1) the supplemental services ordinarily available to students in need of additional reading support referred to as the counterfactual; and (2) the standard ELA courses all students receive. In both Year 1 and Year 2, none of the five high schools participating in the Striving Readers study had a comprehensive approach to address the needs of struggling readers.[7] There was little change in the ELA curriculum from Year 1 to Year 2 in Chicopee; in Springfield changes were implemented to better serve students and meet state standards.

As reported in the Year 1 executive summary, there were planned differences in the dosage of standard ELA based on several factors: each districts' scheduling, supplemental reading supports inclusive or exclusive to ELA instructional time, and graduation and course requirements. There were also *unplanned* differences noted in ELA dosage for three of the five schools, where the total length of the course per day was not delivered as anticipated resulting in differences in the total average amount or dose delivered. Observed differences related to school schedule restrictions were district-specific. Interviews and observations pointed to a great deal of variation in control classroom size.

---

[6] The one teacher who demonstrated no evidence of professional development in Year 2 contributed heavily to the lower overall input score. The absence of professional development for this teacher appeared to be the result of being hired late in the year. Much of the training for this new teacher was designed as "catch-up." In addition, one of the three components contributing to the overall input score (materials/technology) had less consistent ratings for the reasons previously explained.

[7] Students classified as SPED or ELLs had the most access to additional literacy support outside of standard ELA classes. In the absence of such designation, however, the availability of supplemental supports for students was minimal.

There was no evidence of the contamination of control classrooms in either district in Years 1 or 2, which was defined as an infusion of targeted materials or instructional strategies. In addition, the unique characteristics of these interventions were not found to be incorporated in the supplemental services control students received.

Targeted Impacts

Sample Size. In the 2006-07 school year (Year 1), 334 incoming ninth-graders were identified as eligible to participate and were randomly assigned across the five schools. Of these 334 students, 285 (85%) were eligible for placement in the fall: 100 students were assigned to the control condition, 95 to the READ 180 condition, and 90 to the Xtreme Reading condition. In the 2007-08 school year (ear 2), 300 incoming ninth-graders were identified as eligible to participate and were randomly assigned across the five schools. Of these 300 students, 264 (88%) were eligible for placement in the fall: 89 students were assigned to the control condition, 91 to the READ 180 condition, and 84 to the Xtreme Reading condition. The total number of incoming-eligible ninth-graders across the two years was 634; the number eligible for placement was 549 (87%). A total of 16 teachers were assigned in Year 1; 15 teachers participated in Year 2, the majority of these teachers were new.

Analysis Sample. The final number of students in the intent-to-treat or ITT sample was 437; of those, 347 had both a pretest and posttest and were included in the analyses.[8] The percentage of students attaining grade level reading expectations was 12% for both Control and Xtreme Reading students and 14% for READ 180 students.[9] Despite increases in average grade level equivalency or GLE reading scores, there were no statistically significant differences observed between the treatment groups and the control group.

[8] Very few cases missing post-test, approximately 20%, and analysis results were the same when those missing cases were imputed and included.

[9] The percentage of students reading *below* Striving Readers eligibility was high (as measured by the outcome, a different assessment from the screening assessment). However, the percentage of students with reading skills below the fourth grade level was higher in the control group at 42% as compared to READ 180 at 35% and Xtreme Reading at 38%. It is unclear why there would be so many students reading below the targeted level unless this is indicative of difficulties in placement within the first cohort or the reliability of the SRI.

There were no observed (significant) effects of the interventions on student reading achievement, on average, as compared to the control group. That is, no mean difference between treatment and control significant at the $p<.05$ level were observed.

Though statistically non-significant, there were increases in student reading achievement in both treatment groups: average SDRT-4 scores of 665 (READ 180) and 666 (Xtreme) as compared to 662 (Control).[10] The effect size[11] of READ 180 on student achievement as measured was .11; the effect size of Xtreme was .16. When achievement gains are assessed across grade levels, effect sizes have been found to decrease in the upper grades (Bloom, Hill, Rebeck Black, & Lipsey, 2006). Therefore, striving readers in the high schools would generally be expected to gain less than those in the lower grades simply as a result of the trajectory of student growth or development of reading skills.

The performance of interventions *within* each school was similar relative to the control group. That is, student scores generally increased for both groups of intervention participants or decreased for both groups of intervention participants (with the exception of the single school with inexplicably high ratings overall as compared to the population at this school). There were cohort differences overall with higher outcome scores in the first cohort as compared to the second, however they were higher for *the treatment and control groups alike*.[12] Given the challenges with screening and placement in the first year of implementation, this decrease in outcome scores may be most reflective of increased accuracy in the screening process. This hypothesis has some support from the interview data regarding students taking the tests seriously, from teachers and test administrators communicating the importance of the screening test, and from the systems being in place to verify appropriate placements were made for the second cohort. Data from future cohorts will provide a more complete picture of implementation post the first year.

[10] There were two schools for which outcome scores remained high in both cohorts. One school in particular scored unusually high in this sample (especially given the comparison to the population of the school), and the patterns for the remaining four are split relative to the direction of the effect. The two schools with very high control group scores were also the schools with higher overall outcome scores in both cohorts.

[11] Effect sizes were calculated (Glasses) for unadjusted means using the control group standard deviation.

[12] An assessment of treatment and school interactions indicated no significance at the $p<.05$ level (nor were models tested via regression using MCAS as a pretest covariate).

Despite the many complications related to implementation, particularly in Year 1, a pattern of higher targeted implementation levels and higher overall student reading scores was observed in exploratory analyses. While neither intervention's observed difference in impact scores as compared to the control group was significant, the overall sample size was small. Two additional cohorts will participate in this effectiveness trial which will increase and potentially double the current sample size to one at which statistically significant differences may be discernible.[13] The descriptive results discussed here may foreshadow the potential for detecting meaningful intervention effects given increased sample sizes and increased levels of classroom implementation.

Although analyses were conducted for both years combined, implementation levels and impact results varied by year, which itself has implications and at a minimum requires caution when interpreting the findings. It is difficult at this juncture to disentangle the possible influences of newly hired teachers from those associated with changes in scoring specificity (collectively these two influences appear to be positive for READ 180 but negative for Xtreme).

Whole School Evaluation Design Overview

The whole school (SIM-CERT) studies were designed for purposes similar to those of the targeted studies: to assess whether or not participation in the whole-school literacy intervention is associated with overall reading achievement increases over time, and to describe the context in which observed changes occur. Implementation levels were established to characterize the implementation context, and will be examined over time. The evaluation of the school-wide intervention will utilize a quasi-experimental design. Specifically, an interrupted time series analysis will be used to investigate observed changes in the average level of student achievement, over time that may be attributed to the school-wide intervention.

[13] Significance was not tested within individual schools given the sample size and a lack of power to detect within school effects.

Whole School Implementation Study

As with the targeted interventions, scores assessing fidelity to the model were assigned for the whole school literacy intervention. These scores were based on meeting the minimum requirements for two components of implementation: (1) the provision and receipt of professional development and (2) the implementation of SIM-CERT routines in the classroom.

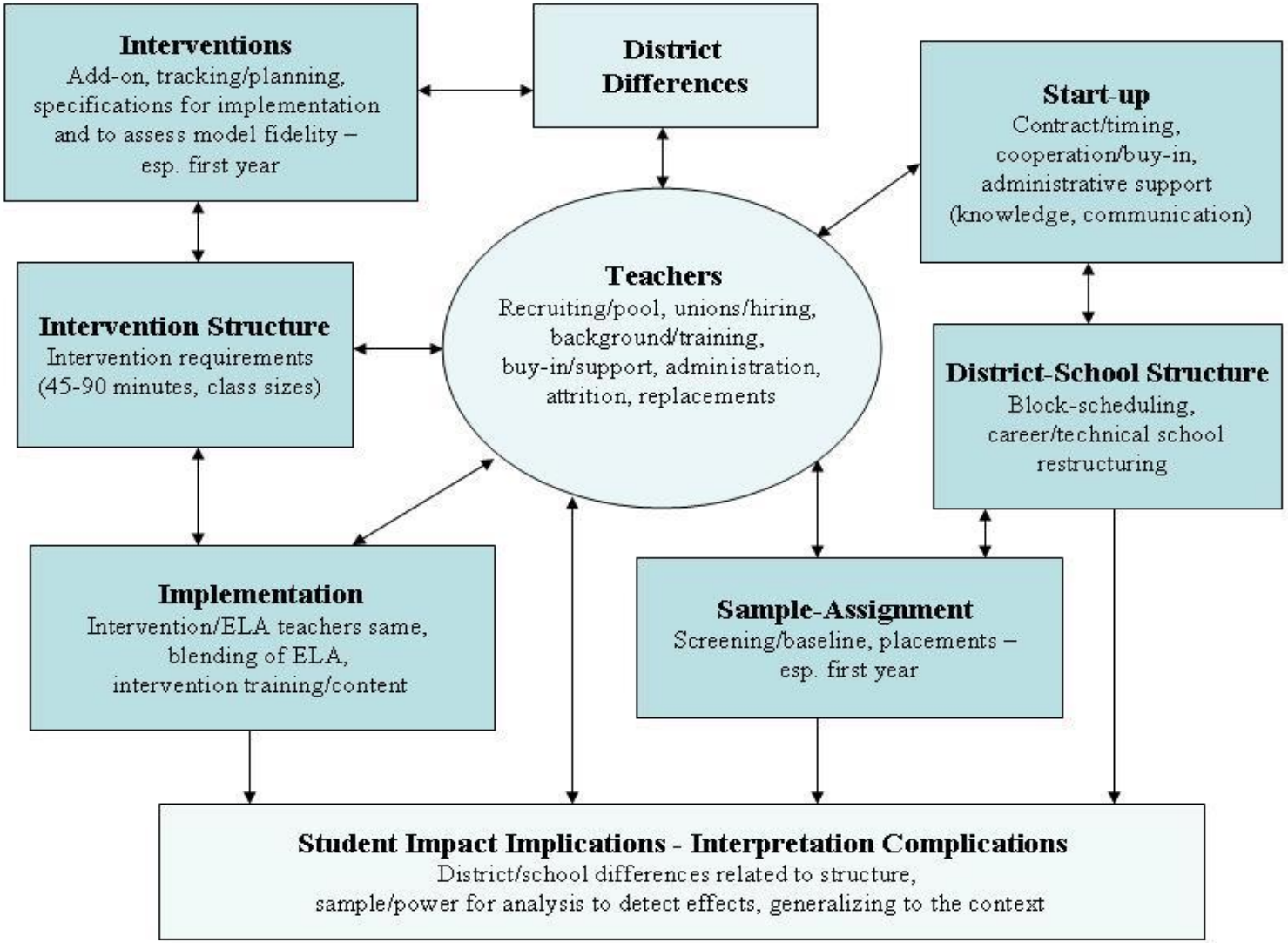
SIM-CERT Implementation. According to model specifications, two initial and two ongoing training sessions were required for teachers during their first year of teaching SIM-CERT. During the second year of participation, teachers were to receive two additional ongoing training sessions.[14] In Year 1 and Year 2, the majority of teachers designated to receive SIM-CERT training across districts and cohorts participated in SIM-CERT initial professional development. Participant numbers increased in Year 2 sessions as compared to Year 1 (in both districts) and the ratings for initial professional development training were relatively high across the two districts. In Chicopee, 98% of Cohort 1 teachers and 100% of Cohort 2 attended the initial professional development. In Springfield, 87% of Cohort 1 and 96% of Cohort 2 teachers attended initial sessions. The proportion of teachers from both cohorts who attended ongoing training during the 1st year of implementation as specified by developers was lower than that attending initial training, particularly in Springfield. None of the Cohort 2 teachers and one Cohort 1 teacher in Springfield received ongoing training as planned before the close of the academic school year. In Chicopee, 71% of Cohort 1 teachers and 62% of Cohort 2 teachers attended ongoing training as planned. The majority of teachers (Cohort 1 only) received ongoing professional development during their 2nd year of implementation: 74% in Springfield and 92% in Chicopee.

The majority of SIM-CERT teachers, across cohorts and districts, met minimum developer-defined requirements for classroom-level implementation in Year 2; that is, most teachers reported use of the Unit Organizer routine as well as one other routine during the 2007-2008 academic year. Levels of implementation were higher in Chicopee than in Springfield (as reported by teachers, literacy coaches, and administrators).

[14] Teachers were to receive ongoing training during the school year in which they were expected to implement the routines in the classroom.

Barriers to Implementation

Several reported barriers were responsible for variability in the degree of implementation of both the targeted and whole school interventions and necessitated adaptations to the planned models. Barriers often emerged and sometimes mushroomed as a result of complicated circumstances. While some barriers were caused by school-related structural and organizational (e.g., scheduling) issues, others were created as a result of the intervention requirements. Despite these difficult circumstances, district and Striving Readers' staff continually made adjustments that were designed to eliminate or at least diminish the impact of as many barriers to implementation, and the research study, as possible. The following graphic illustrates a majority of the barriers and their inter-connections.


Interventions
Add-on, tracking/planning, specifications for implementation and to assess model fidelity – esp. first year
District Differences
Start-up
Contract/timing, cooperation/buy-in, administrative support (knowledge, communication)
Teachers
Recruiting/pool, unions/hiring, background/training, buy-in/support, administration, attrition, replacements
Intervention Structure
Intervention requirements (45-90 minutes, class sizes)
District-School Structure
Block-scheduling, career/technical school restructuring
Implementation
Intervention/ELA teachers same, blending of ELA, intervention training/content
Sample-Assignment
Screening/baseline, placements – esp. first year
Student Impact Implications - Interpretation Complications
District/school differences related to structure, sample/power for analysis to detect effects, generalizing to the context

A few additional barriers specific to particular interventions should also be noted. Changes by SIM to the Xtreme model for example, while in response to teacher's comments about the organization of materials in Year 1, caused confusion among teachers and may have resulted in a less than optimal implementation of the intervention. Moreover, teacher attrition posed particular challenges for the whole school intervention. Since SIM-CERT is, by design, a school-wide intervention, higher than expected administrator and SIM-CERT teacher turnover could affect teacher buy-in and the spread of the intervention, thereby diminishing the possibility of positively influencing student-literacy outcomes. Finally, teachers in both targeted interventions identified classroom management issues, mandatory testing, competing district initiatives and perceptions of student misplacement as key barriers to their ability to deliver instruction according to model requirements. Developers noted variation in teaching skills, especially in the first year of implementation, as a challenge faced as well as determining the best way in which to support individual teacher and district needs.

Evaluation Summary

The Springfield and Chicopee school districts have overcome many obstacles in the development, planning, and implementation of their Striving Readers (SR) grant. In particular, two dissimilar districts have implemented two targeted interventions (all other SR grantees implemented only one) as well as one whole school intervention.

Many of the barriers presented in the implementation of the grant in the first year resulted from both contextual and contractual factors, which did not necessarily emerge from the intervention models themselves but may have resulted from attempts to fit the models as required into this context (refer to the logic models for an overview of context). Some of the contextual factors included: the urban setting, population, and student needs; the various policies of the schools and districts addressing scheduling, administrative issues, etc.; as well as general staffing and personnel matters.[15]

[15] One of the districts SR program leads took another position elsewhere prior to the first school year of grant implementation.

Contractual complexities specifically refer to the requirements for the grant implementation; the monitoring and oversight of the fidelity of implementation; and the observance of the rigorous research specifications.

Given the challenges inherent in both creating a successful collaboration between two districts and implementing two interventions, it is not surprising that complexities arose which would not normally be encountered in a standard literacy program implementation. An initial barrier related to the rigorous research requirements, for example, involved the cooperation, ability, and willingness of both districts to incorporate a "true" control group to address the counterfactual (i.e., *what would happen in the absence of treatment*). Additional challenges involved the need to standardize implementation across two very different district and school systems.

Intervention plans necessitated consistent tailoring to accommodate rigorous research study requirements and unanticipated time by district staff and evaluators was spent to ensure successful implementation. At the same time, districts faced changes in lead program staff, challenges related to communication with stakeholders and participants, and complications in screening and placing the population of students who were randomly assigned to participate in the targeted interventions.

Progress was made in overcoming these barriers, particularly in Year 2. Districts implemented each of the targeted interventions while maintaining the integrity of the randomized controlled trial design and assignment and repeatedly demonstrated their commitment to ensuring the success of the grant. District staff collaborated fully with evaluators in all phases of the evaluation. Their serious consideration of any potential positive or negative influences on study outcomes as well as "full disclosure" has been commendable. Such diligence ensures that the final results of this study will produce information that can be used by policymakers, district administrators, and school staff to make confident choices regarding effective interventions for their students.

I. Introduction and Study Background

This report presents implementation and impact findings based on district documentation and data gathered by The Education Alliance from the second year (2007-08) of the Striving Readers grant as implemented by the Springfield and Chicopee Public School Districts. The Striving Readers grant requires the implementation of both targeted and whole-school literacy interventions. In the Springfield and Chicopee Public School Districts, five high schools (three in Springfield and two in Chicopee) are implementing two targeted interventions—both developed using scientifically-based research (SBR) to promote the reading skills of struggling readers—as well as a whole-school intervention developed to promote reading skills throughout the student population.

The targeted interventions are: (1) READ 180 Enterprise Edition (Scholastic, Inc.) and (2) Strategic Instruction Model (SIM) Xtreme Reading (University of Kansas, Center for Research on Learning). Both targeted interventions have been provided as a supplement to the regular English Language Arts curriculum in the participating schools. The whole school intervention is the Strategic Instruction Model Content Enhancement Routines for Teachers (SIM-CERT), which along with Xtreme Reading is a part of the University of Kansas' Content Literacy Continuum (University of Kansas, Center for Research on Learning).

The U.S. Department of Education (ED) and its contracted Striving Readers technical assistance provider, Abt Associates, have made significant contributions to this report.

II. District Context

Springfield, Massachusetts: Background

The Community. According to the 2006 U.S. Census, the mid-sized city of Springfield is a community of 152,082 people, located in western Massachusetts. Twenty-nine percent of Springfield's population comprises children under the age of eighteen. Approximately 23% of the overall population and more than 75% of all public school students in Springfield live in households at or below the poverty line.[16]

Springfield Public Schools (SPS). Springfield Public Schools enrolled approximately 25,233 students in the 2007-08 school year.[17] Springfield is the second largest school system and one of the lowest performing school districts in the state. Springfield has four high schools, three of which are participating in the Striving Readers Program.[18] The district is a Title I District and although the three high schools—High School of Commerce, Putnam Vocational-Technical High School, and the Springfield High School of Science and Technology (SciTech)—are non-Title I schools by designation, they qualify as schools eligible to receive Title 1 funds (MADOE, 2007).[19] Additionally, all three high schools participate in the Metropolitan Council for Educational Opportunity (METCO), a state-funded program designed to address racial imbalances by busing children from urban areas to surrounding suburban areas (Metropolitan Council for Educational Opportunity, n.d.).

[16] Local poverty statistics obtained from a district document downloaded from www.sps.springfield.ma.us, November 7, 2007.
[17] Data were obtained from the Massachusetts Department of Education's District Profiles database, http://profiles.doe.mass.edu/, January 5, 2009.
[18] This does not include the numerous alternative secondary schools and private secondary schools located in Springfield.
[19] This is true of Chicopee High Schools as well. Eligibility for the high schools relies upon what one Striving Readers program manager referred to as a "calculation of preponderance"—meaning that although the number of students in the high schools registered for free or reduced lunch does not necessarily reflect a percentage that warrants Title I status, the preponderance of other factors (most notably the Title I status of all sending middle schools) indicates that the actual number of known free/reduced lunches in the high schools is lower than the actual number of students qualifying.

A state-appointed financial control board currently governs Springfield's public schools as well as the City of Springfield. The dire financial status of the city and the district, in addition to past teacher contract difficulties, has contributed to significant losses of teachers, other personnel, and services to the public schools.

Chicopee, Massachusetts: Background

The Community. A neighboring community of Springfield, Chicopee has 23,117 households where 23% percent of the population comprises children under the age of 18. The median household income is $35,672 and approximately 12% of the overall population lives below the poverty line (U.S. Census, 2006).

Chicopee Public Schools (CPS). Chicopee has two high schools, both of which are participating in the Striving Readers Program. Like Springfield, Chicopee is a Title I District with its two participating high schools eligible to receive Title I funds. Chicopee also participates in the METCO Program. Chicopee Public Schools enrolled 7,754 students in the 2007-08 school year (MADOE, 2009).

Characteristics of Participating Schools

Descriptive information for every high school participating in the Striving Readers Program is presented in Exhibit 1.

Exhibit 1. Characteristics of participating schools, 2007-08 [20]

Characteristics	Chicopee Schools		Springfield Schools			State
	CHS %	CCHS %	Putnam %	SciTech %	Commerce %	
Non-White	27	20	87	85	91	29
First Language Not English	15	9	25	28	26	15
Limited English Proficient (LEP)	2	1	10	17	10	6
Low Income	42	35	69	67	73	30
Special Education	15	13	22	27	24	17
Total Number of Students	**1225**	**1320**	**1472**	**1622**	**1445**	**--**

Source: Massachusetts Department of Education. *School/District Profiles.* Retrieved November 12, 2008 from http://profiles.doe.mass.edu/

Graduation Requirements and Adequate Yearly Progress (AYP) Status

The five Springfield and Chicopee high schools operate in a high-stakes climate with strict, state-mandated graduation requirements. In the 2007-08 school year, high school students, over 60% of whom are from minority populations (88% in Springfield), and over 58% of whom live in poverty (70% in Springfield), must have passed the Massachusetts Comprehensive Assessment System (MCAS) in order to graduate from high school.

As required by the federal No Child Left Behind Act (NCLB), all schools and districts are expected to meet or exceed specific student performance standards in English Language Arts/Reading (ELA) by the year 2014. In order to monitor progress toward set performance goals, state departments of education issue adequate yearly progress (AYP) determinations yearly. Exhibit 2 depicts the performance history of the Springfield and Chicopee districts in ELA by providing a snapshot of AYP status for 2006 (the year of the grant application), as well as 2007 and 2008 (the first and second year of implementation of the Striving Readers Program).

[20] The characteristics of the participating schools were similar to those reported for the 2006-07 school year in the Year 1 report, so are only reported for Year 2.

Exhibit 2. AYP determination for ELA by district (2006, 2007, and 2008)

	Chicopee			Springfield		
	2006	**2007**	**2008**	**2006**	**2007**	**2008**
Grade Span 6-8						
Aggregate	Not met	***Met AYP Criteria***	Not met	Not met	Not met	Not met
Subgroup	Not met	Not met	Not met	Not met	Not met	Not met
Grade Span 9-12						
Aggregate	Not met	***Met AYP Criteria***	***Met AYP Criteria***	Not met	Not met	Not met
Subgroup	Not met	Not met	Not met	Not met	Not met	Not met

Source: Massachusetts Department of Education, *School and District Accountability*. Retrieved on 10/30/2008 from http://www.doe.mass.edu/sda/

The trend of district accountability data demonstrates the need for literacy support for both middle school and high school students. In Chicopee, at the high school level, aggregate scores met AYP criteria for the past two years, but subgroups continued to lag behind. In 2008, subgroups that did not make AYP include special education, low income, and Hispanic/Latino students. In Springfield, AYP criteria have not yet been met. Subgroups in Springfield high schools that continue not to meet AYP benchmarks include African American, low income, Hispanic/Latino, special education, and Limited English Proficient (LEP) students. As stated in the previous report, the fact that these subgroups are not making AYP is particularly relevant given that a majority of students (more than 50% aggregated across both districts) in the participating high schools are either African American, Hispanic/Latino, or are living in poverty.

The accountability status for each of the five participating schools is presented in the exhibit below. In 2008, the Chicopee schools were designated as "Improvement Year 1" after two consecutive years of not making AYP requirements. In such cases, the Massachusetts accountability system requires that the schools offer parents the option of sending their child to another school within the district that has made AYP if space in one is available.

Exhibit 3. ELA accountability status for participating high schools (2008)

District	School Name	2008 ELA Status
Chicopee	Chicopee Comprehensive High School	Improvement Year 1 - Subgroups
Chicopee	Chicopee High School	Improvement Year 1 - Subgroups
Springfield	High School Of Commerce	Restructuring Year 2+ - Aggregate
Springfield	High School of Science and Technology	Restructuring Year 2+ - Aggregate
Springfield	Putnam Vocational-Technical High School	Restructuring Year 2+ - Subgroups

Source: Massachusetts Department of Education, *School and District Accountability*. Retrieved on 10/30/2008 from http://www.doe.mass.edu/sda/

In 2008, the Springfield schools were designated as "Restructuring Year 2" status.[21] The district's only Chapter 74 approved vocational-technical program school was also designated by the state as "chronically under-performing." This school was offered the opportunity to convert to a Commonwealth Pilot School by the Massachusetts Board of Elementary and Secondary Education (the Board) in November 2006—the first year of implementation of the Striving Readers Grant. The Commonwealth Pilot School initiative was intended to introduce substantive reform to schools struggling with low student achievement by providing increased flexibility to organize schools and staffing to best meet students' needs, while operating within the economies of scale of a larger public school district. Commonwealth Pilot Schools are unique in the nation in that, by virtue of an innovative teacher union contract, such schools have autonomy in five operational areas: (1) budget, (2) staffing and hiring, (3) governance, (4) curriculum and assessment, and (5) the school schedule and calendar. The vocational-technical school converted to Commonwealth Pilot School status. The vocational-technical school was also granted autonomy to create smaller schools within the original school and currently, there are five smaller schools within the vocational-technical school. In June 2008, this school concluded its first year of operation as a Commonwealth Pilot School.

[21] Description of MADOE's accountability system was downloaded from a recent news brief posted on the MADOE website http://www.doe.mass.edu/sda/news08/pr091908.html on 10/28/08.

III. Theoretical Rationale and Description of Interventions

Two targeted interventions (READ 180 and Xtreme Reading) were selected by the Springfield-Chicopee[22] school districts to improve the reading skills of struggling readers. Both READ 180 and Xtreme Reading were to be implemented as "add-on" or supplemental interventions. That is, the interventions were implemented in addition to the regular English Language Arts (ELA) class required in the participating schools.[23] The whole school intervention model, SIM-CERT, was selected to improve literacy across content areas and its implementation was to be phased in over the period of the grant.

The following descriptions summarize key elements of the interventions, as planned and as implemented, and any changes occurring in each year and over time.

READ 180 Targeted Intervention

The READ 180 program is an intensive literacy curriculum developed for struggling readers in grades 4 through 12 to bring their reading skills to grade level standards and to promote reading comprehension. Initially developed in 1985 by Ted Hasselbring at Vanderbilt University, the program, then named the Peabody Literacy Lab, uses anchored instruction (Hasselbring & Goin, 2004). Anchored instruction is based on a philosophy of using authentic situations as anchors to "enable students to practice noticing and resolving problem situations" (p.138). The READ 180 program also uses computer-assisted instructional (CAI) software to track individual student progress and to adjust reading instruction accordingly.

[22] Springfield-Chicopee is used as an abbreviation for the Springfield Public Schools and Chicopee Public Schools implementing their jointly proposed Striving Readers program.

[23] Those students wanting to take an elective such as art, for example, needed to wait until the upper grades to take it. Physical education, which is not an elective but is required for one semester per year, was perhaps doubled-up in upper grades to fulfill this requirement.

Using the concept of anchored instruction, the CAI software has "an animated tutor who guides the student and provides feedback via a digitized human voice" (p.133). After purchasing the rights to the Peabody Literacy Lab Program and changing its name to READ 180, Scholastic contributed significantly to the program's further development (Scholastic, Inc., 2005a).

The goal of READ 180 is to help struggling adolescent readers achieve proficiency in reading at grade level. Objectives of the program include targeting specific elements of phonics, fluency, vocabulary, comprehension, spelling, writing, and grammar, and promoting self-directed learning (Scholastic, Inc., 2005c). READ 180 materials are written specifically for adolescents. The stories contain content that is of interest to their particular age group and is connected to their everyday experiences.

READ 180: Instructional Approach and Curriculum

The READ 180 instructional model provides structure to classroom activity, and is based on a 90-minute block that begins and ends with whole-class instruction directed by the teacher (see Exhibit 2). The model begins with 20 minutes of whole-class instruction in which skills are explicitly taught in the areas of word analysis, vocabulary, and reading comprehension, and concludes with a 10-minute whole-class wrap-up (Scholastic, Inc., 2005a). For the 60 minutes in between whole-class sessions, students break out into smaller groups and rotate among the following three stations:

(1) Small-group direct instruction through which the teacher focuses on needs specific to the selected group of students;
(2) Independent student work using READ 180's CAI software; and
(3) Modeled or independent reading from paperbacks and/or audio books.

Not only does the READ 180 intervention provide classroom structure, it also provides content through specific teacher resources (e.g., rBook Teacher's Edition, Anchor videos) and student materials that are used during the whole-class and small-group sessions.

The rBook Teacher's Edition provides teachers with content and instructional routines to further develop students' reading comprehension, vocabulary, writing, and grammar skills as well as encourage active participation.[24] Anchor videos used during the whole-class direct instruction segment of the class, on the other hand, are used to jump-start the whole-class activity. They provide background knowledge and are designed to capture student interest by raising provocative questions. The rBook's nine workshops are estimated to require one school-year (approximately eight months) of instruction, as per the READ 180: Stage C rBook Teacher's Edition (2005)—a manual that contains explicit teaching instructions for using the rBook for whole-class and small-group instruction. While teachers have the rBook Teacher's Edition, students are provided with their own rBooks that are used as interactive work texts.

Teachers are to use specific READ 180 instructional strategies during READ 180 teacher-directed activities in whole and small groups. For example, in the whole-class segment teachers may:

- use anchor videos and discussions to build background knowledge before reading;
- create opportunities to hear models of fluent reading;
- teach and model reading skills and strategies;
- use explicit instruction of important academic vocabulary words and word study elements;
- provide instruction in key writing types;
- deliver lessons in grammar; and
- teach structured engagement routines that involve students in their learning (i.e. RED Routines).

[24] Instructional routines covered include: teaching vocabulary, oral cloze, think (write)-pair-share, idea wave, numbered heads, the writing process, and peer feedback.

In small-group segments, teachers may use many of the whole-class strategies, and may also offer differentiated instruction in phonics, fluency, vocabulary, word study, spelling, and comprehension. They may also provide fluency assessment and practice, and conduct teacher conferences to set goals, check reports, reflect on books, and review rBooks.[25]

The READ 180 developers recognize that professional development must be ongoing for teachers to improve their instructional strategies and techniques in a manner that ultimately improves student literacy. READ 180's professional development is designed "to help teachers be successful and to foster and sustain best teaching practices in the classroom" (Scholastic communication, 2007). Accordingly, READ 180 offers a variety of professional development opportunities and support, ranging from trainings, seminars, in-classroom support, web-based instructional support and online RED courses focused on aspects of reading instruction. READ 180's professional development model is illustrated in the logic model on the following page. The professional development provided to teachers is one of the "inputs" or resources provided (in addition to materials and support) in order to reach the ultimate goal of improving student outcomes.

A logic model depicting the key components of the READ 180 intervention (as planned and expected outcomes) is depicted in Exhibit 4. This model reflects changes as planned in Year 2.

READ 180: Year 1 and Year 2 Changes

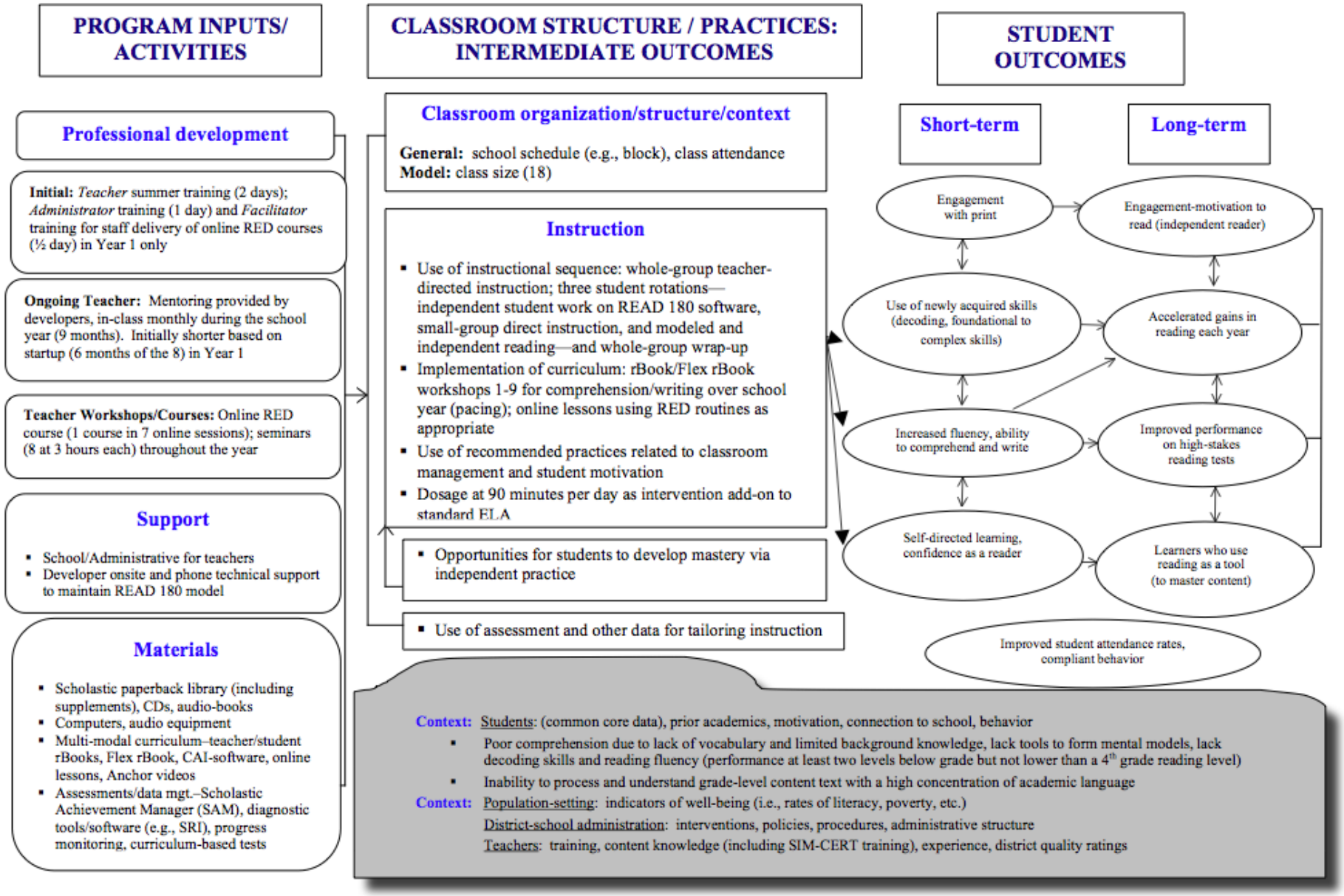
In Year 2, as per the Striving Readers (SR) district implementation team, most of the changes from Year 1 involved the delivery of professional development. More specifically, in Year 2, the plan did not include administrator training at the beginning of the school year as key administrators were familiar with the intervention after participation and because a prior version of READ 180 had already been used in Springfield schools. The facilitator for the online RED course, offered to all participants, was trained in Year 1. As a result, no additional facilitator training was required in Year 2.

[25] Teacher Implementation Guide, p. 36.

There were nine monthly in-class teacher visits planned for the second year which differed from the six possible (and so planned) in the first year during start-up.[26] Finally, the number of planned seminars for teachers stayed the same from year to year as did plans for one online RED course. However, the number of online sessions comprising the single course had not been specified in Year 1 within the logic model (in Year 2, it was determined that there were a total of seven).

[26] Note that Scholastic indicated monthly visits began in October based on later communications and, therefore, at least eight should have been conducted. However, this differed from the Striving Readers implementation team reported visits as well as the number reported by teachers. A more conservative estimate of planned monthly visits was used in the first year given the complications schools faced in scheduling and the setup of needed technology. Otherwise, input scores (for assessing implementation fidelity) would have been adversely affected for READ 180.

Exhibit 4. READ 180 logic model



As planned, students who received one year of READ 180 in 2006-07 but did not have outcome test scores (SDRT-4) that met grade level expectations were to be provided with a second year of READ 180. These students worked from the already-developed Flex rBook that parallels the content of the rBook (the student book designed to be a resource for students in whole-class and small-group instruction) without duplicating the same texts.[27] Although there was a review of the same skills in the second year, including summarizing for comprehension, teachers were to use differentiation to address student needs and to increase the level of sophistication of the skills learned so that these literacy skills could be applied to different content areas/subjects.[28] Additionally, per the Striving Readers district implementation team, more complex texts were introduced to the students receiving a second year of READ 180. Developers provided books with more challenging reading for those at higher levels as well as additional titles at the lower Lexile levels for greater variety.

Xtreme Reading Targeted Intervention

The Xtreme Reading Program of the Strategic Instruction Model (SIM) was developed by the University of Kansas Center for Research on Learning (KU-CRL). While READ 180 focuses on the fundamentals of reading, Xtreme Reading has a meta-cognitive approach focusing heavily on explicit strategy instruction.

Xtreme Reading is part of the Content Literacy Continuum, a framework of literacy supports that vary in intensity depending on student need. The KU-CRL model is based on research indicating that content literacy occurs not only when students have mastered the critical content as determined by teachers, but also when students can manipulate and generalize this content to other learning situations. This framework of adolescent literacy support is based on research that, in order to thrive throughout their academic careers, adolescents must be able to read and understand large volumes of complex and difficult reading materials.

[27] These texts are not sequential, so a whole class may start in either the rBook or the Flex rBook and then alternate to the other text the following year, when needed.
[28] Information provided by Karen Burke, Scholastic, November 2008.

Xtreme Reading is the name associated with Level 3 of the Content Literacy Continuum and was developed for adolescents who struggle with reading and writing. More specifically, Xtreme Reading targets students reading at least two years below grade level but who read at or above the fourth grade level. Xtreme Reading focuses on intensive strategy instruction, particularly reading instruction that helps students to develop accurate word recognition and increased fluency and comprehension. The program addresses the skills and strategies needed to bring meaning to reading so students will learn how to read at grade level. The SIM approach to instruction as described by KU-CRL involves intensive, carefully tailored lessons in which students have abundant opportunities to practice targeted learning strategies that will help them succeed in their classes.

The team from KU-CRL trains teachers in all aspects of what are called "Learning Strategies" for students. The professional development model for KU-CRL includes initial training, ongoing in-class mentoring by providers, as well as additional workshops on specific routines. These strategies prompt teachers to organize, clarify, and standardize student approaches to engaging with and mastering content. The Learning Strategies (Level 3) combined with the SIM-CERT or Content Enhancement Routines for Teachers (Levels 1 and 2) comprise the three levels of the Content Literacy Continuum implemented in this program (refer to Exhibit 5).

Exhibit 5. SIM Content Literacy Continuum (CLC)

Level	Purpose	Instruction
1	Master critical content	Enhanced content instruction (strategic teaching to ensure mastery of critical content for all students)
2	Use learning strategies across classes	Embedded strategy instruction (teachers embed selected learning strategies in core curriculum courses)
3	Master specific reading strategies (e.g., self-questioning, visual imagery, paraphrasing)	Explicit strategy instruction *(Xtreme Reading)*

Source: Dr. Faddis (personal communication, November 2007), RMC Research Corporation, Portland, Oregon, based on information provided by Susan Robinson, University of Kansas, Center for Research on Learning.

Xtreme Reading: Instructional Approach and Curriculum

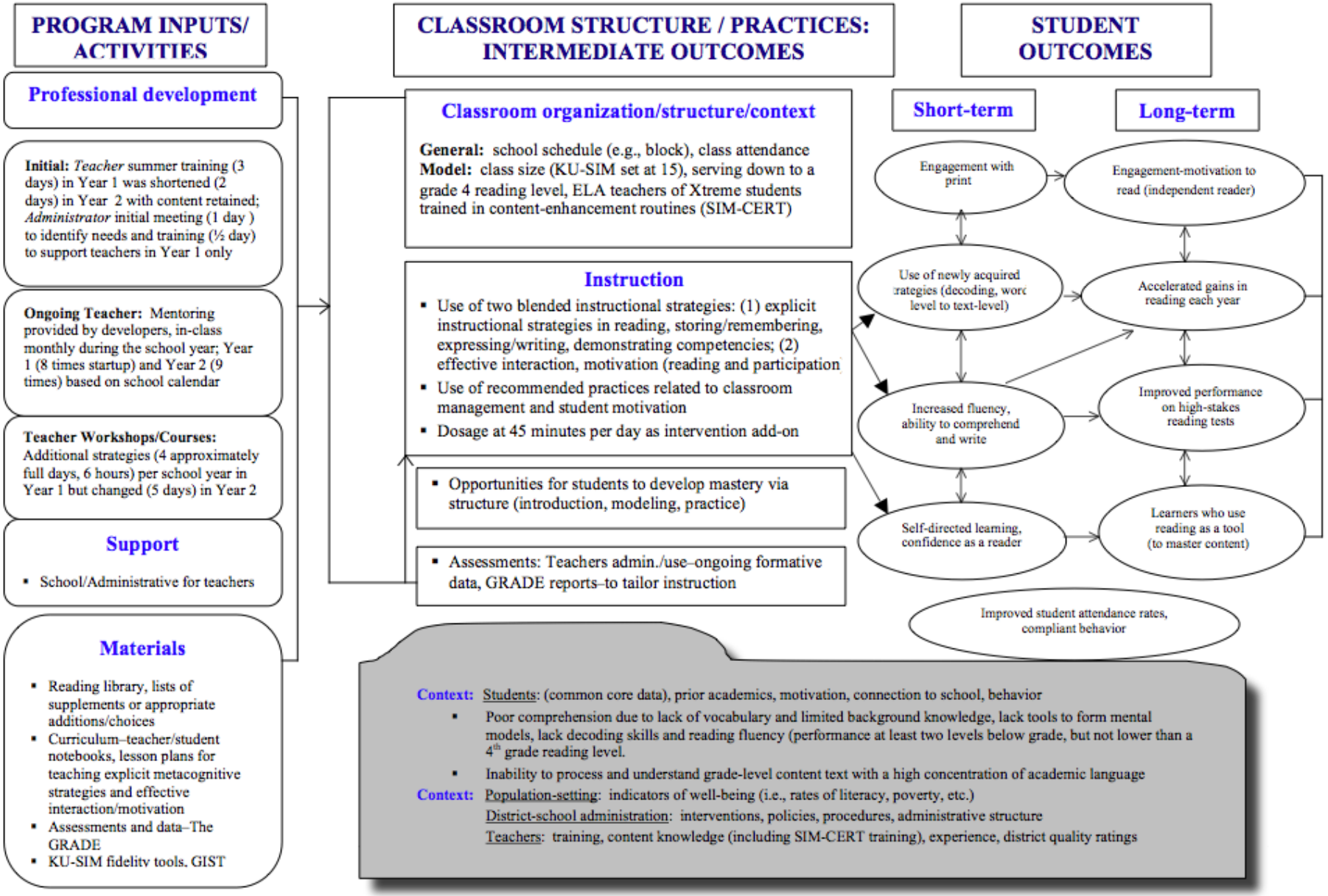
The year begins with units addressing behavior (ACHIEVE, Talking Together, SCORE) and motivation (Possible Selves) in which students learn about what is expected of them in the classroom and how to create a productive learning environment. Students are explicitly taught the appropriate behaviors for specific classroom situations including lectures, discussions, independent work, and small-group work. The Possible Selves unit focuses specifically on student motivation and involves having students analyze their current lives and then set goals to enhance their futures (Xtremereading.com). The behavioral and motivational portion of Xtreme Reading takes approximately four weeks to implement. Note the changes regarding these units in Year 2 below and on the following pages.

The Xtreme Reading program then shifts to the seven reading strategies: LINCS Vocabulary, Word Mapping, Word Identification, Self-Questioning, Visual Imagery, Paraphrasing, and Inference. The first three strategies focus on vocabulary development (although the LINCS model focuses on learning meaning of new words through memorization, as well as on advanced phonics and decoding for multi-syllabic words) while the remaining four strategies target reading comprehension. Reading comprehension is addressed using strategies such as imagery (i.e., teaching students to create mental pictures as they read), paraphrasing (i.e., teaching students to identify the main points of a paragraphs and then restate it in their own words), as well as prediction and questioning. The program also encourages teachers to support reading fluency through explicit teaching and modeling for students. In addition to the reading strategies, Xtreme Reading integrates writing strategies (such as Paragraph Writing and Theme Writing) with reading instruction. These writing strategies focus on the writing process and thus emphasize planning, writing, providing or accepting feedback, and editing (source: Xtremereading.com).

The Xtreme Reading model uses an instructional approach that involves both teacher-directed whole-group discussions, teacher modeling of strategies, guided practice activities, and paired-student practice as well as independent practice.

Xtreme Reading teachers receive direct training in the Learning Strategies and SIM-CERT strategies as well as ongoing consultation services from the SIM developers (i.e., KU-CRL staff). Xtreme Reading instructional strategies fall into six categories: (1) reading, (2) storing and remembering information, (3) expressing information (writing), (4) demonstrating competence, (5) effectively interacting with others, and (6) motivation. These strategies include components of reading as well as class participation. A logic model depicting the key components of the Xtreme Reading intervention as planned and expected outcomes are depicted in Exhibit 6, which includes any Year 2 changes as planned.

Exhibit 6. Xtreme logic model



Xtreme Reading: Year 1 and Year 2 Changes

The changes in Xtreme Reading from Year 1 to Year 2 mostly involve changes to the professional development model, changes to Xtreme Reading materials, and the development of "Strategic," which is the version of Xtreme Reading to be delivered to students needing a second year of the targeted intervention.

In Year 2, as per the SR district implementation team, the plan did not include administrator training at the beginning of the school year. As reported in the Annual Performance Report, administrators met with developers and the Striving Readers work group to learn model requirements and to discuss how to support implementation in their district.[29] As for initial teacher professional development, two days were planned as approved by the developer in Year 2, whereas three days of initial training were planned in Year 1. The SR district implementation team reported that the content and topics covered in Xtreme Reading initial trainings was the same across both years but "condensed."

Additionally, per the SR district implementation team, while Xtreme Reading teachers in Year 1 attended the training in August 2007 for SIM-CERT Cohort 1 teachers, it was later determined that these teachers needed more time for Xtreme Reading training rather than SIM-CERT training. So, Xtreme Reading teachers did not participate in any remaining or subsequent SIM-CERT professional development activities. Developers indicated any necessary SIM-CERT training was embedded in Xtreme Reading sessions or monthly coaching. In terms of ongoing professional development, there were nine monthly in-class teacher visits planned for the second year which differed from the eight possible (and so planned) in the first year during start-up.

Toward the end of Year 2, developers indicated that they modified Xtreme Reading materials and changed the yearly pacing calendar in response to teacher request.[30]

[29] The Striving Readers district implementation team comprises the Chief Implementation Officer, the Chief Communications Officer, the two Striving Readers Specialists (district coordinators), and two Management Support Specialists. This district team meets regularly to work on planning and grant implementation.

[30] SIM developers reiterated that it is an experimental version and revisions have typically been ongoing during the Striving Readers studies.

The initial units on student behavior and motivation were abbreviated and in some cases, only to be covered "as-needed." In addition to changes in the pacing calendar, more titles were provided in the Xtreme Reading library to address higher reading levels as well as to provide more variety in the reading material for students, per the district team. SIM does provide Lexile levels on selections included in the libraries.

Finally, as planned, students who received one year of Xtreme Reading in 2006-07 but did not have outcome test scores (SDRT-4) that met grade level expectations were to be provided with a second year of Xtreme referred to as Strategic—a continuation of Xtreme Reading intervention practices.[31] However, Strategic was not yet developed to the extent necessary so districts collaborated with the SIM team in both years to specify this intervention. There was strong district commitment to developing this second year curriculum with SIM that was based on the awareness that a number of Year 1 students would need another year of targeted intervention.

Whole School Intervention

The Strategic Instruction Model–Content Enhancement Routines for Teachers (referred to throughout as SIM-CERT) comprises reading strategies used to improve literacy instruction across all disciplines. KU-CRL reports developing these based on more than 20 years of reading research. The intervention comprises Levels 1 and 2 of the Content Literacy Continuum (CLC) and is designed to help students understand critical course content (refer to Exhibit 5). The overarching goal of SIM-CERT implementation is to empower teachers to facilitate and students to develop content literacy. Content literacy is defined as the engagement skills and strategies (including listening, speaking, reading, and writing) necessary to process, understand, and master material across a range of academic disciplines.

[31] Strategic was designed to be more fluid and adaptable to the needs of the students than its predecessor and focuses on strengthening student ability to work independently with a broader range of materials and within a variety subject areas; the goal is advancing skills in reading and writing literacy. In Strategic, teachers may re-teach portions of Xtreme Reading, like sentence structure or the prerequisite skills for writing a good paragraph, with the expectation that skills and strategies will apply to a variety of text materials. The larger focus of Strategic is more global, however; its emphasis is on writing strategies, including paragraph writing, theme writing, and integration of strategies across materials and subject matter, which can take place once the prerequisite skills are re-taught and mastered by the students.

SIM-CERT: Instructional Approach and Learning Strategies

Content enhancement begins with the provision of meta-cognitive strategies for teachers to evaluate and therefore improve their practice. The developers of SIM-CERT have identified three key teacher activities to promote content enhancement: evaluate the content, determine the necessary approaches to learning for student success, and teach with routines and instructional supports that assist students as they apply appropriate techniques. By following these steps, teachers will identify and demonstrate to students the goal or product of learning as well as model the method by which learning occurs. Teachers must assess student characteristics such as intellectual curiosity, interest in the subject matter, and general motivation to learn. Teachers must also decide on appropriate and customized instructional strategies or routines. By matching instructional approaches with the learning characteristics of students, teachers can differentiate their instruction to meet individual student needs.

The developers[32] note that the explicit instruction of the strategies used is critical for two reasons. First, specificity is required in order for teachers to impart the details of given approaches to students (and to be sure students understand), and second, because explicit instruction on these approaches or routines teaches students *how* they are learning, in addition to *what* they are learning. There are four categories of strategies, termed Enhancement Routines, to be used by teachers in the following areas: planning and leading learning; exploring text, topics, and details; teaching concepts; and, increasing student performance (refer to Exhibit 7).

Exhibit 7. SIM Content Enhancement Routines for Teaching (SIM-CERT)

Planning and Leading Learning	**Teaching Concepts**
▪ Course Organizer	▪ Concept Mastery Routine
▪ Unit Organizer	▪ Concept Anchoring Routine
▪ Lesson Organizer	▪ Concept Comparison Routine
Exploring Text, Topics, and Details	**Increasing Performance**
▪ Framing Routine	▪ Quality Assignment Routine
▪ Survey Routine	▪ Question Exploration Routine
▪ Clarifying Routine	▪ Recall Enhancement Routine
▪ Order Routine	▪ LINCing Routine

Note: Based on information provided by Dr. Robinson, University of Kansas, Center for Research on Learning, November, 2007 (Source: Dr. Faddis, RMC Research, Portland, Oregon).

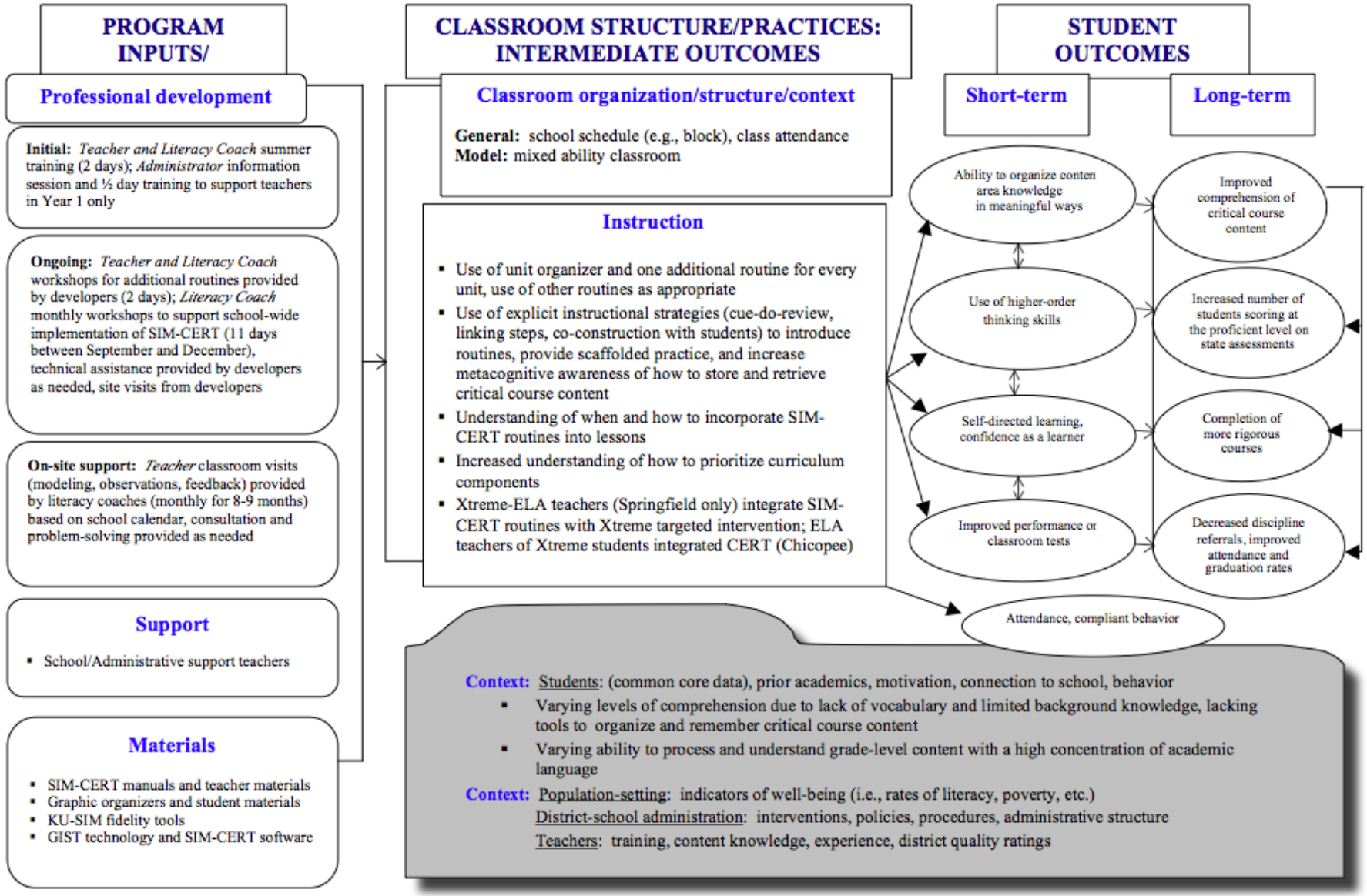
[32] KU-CRL is the developer of SIM (the developers of SIM are referenced throughout this document).

These categories represent the four general task areas that teachers are to engage in as they evaluate, organize, prepare, deliver, and enhance content delivery for students. Each of the four categories or Enhancement Routines has several subcategories. For example, the first category, "Teaching Routines for Planning and Leading Learning," has three "Organizer" subcategories—for the whole Course, Units, and Lessons. The routines teachers use depend on their needs or the needs of their department as well as the relevance of the routine to the content being taught as per the intervention. Ongoing training as planned was designed to provide follow-up support to teachers during the academic school year in which they are implementing the intervention. A critical method of intervention delivery for the SIM-CERT model is the employment of a school-embedded SIM-CERT literacy coach. This coach is trained intensively by the SIM network of trainers to provide ongoing on-site support to teachers.

A nationwide SIM-CERT trainer network, overseen by KU-CRL, works directly with teachers and districts to create opportunities for faculty to learn the SIM-CERT instructional practices as well as to promote and support the use of these strategies in the classroom in a manner that is customized to school needs. Prior to implementation, individual interviews with teachers allow SIM-CERT trainers to gather information about teacher challenges, student needs, and cultural norms specific to the school. Additionally, during interviews trainers explain the content and process comprising upcoming trainings. Moreover, individual interview information is used to develop vignettes and themes for whole-class training.

Exhibit 8 features a logic model that depicts the key components of the SIM-CERT intervention (as planned and expected outcomes), which includes any Year 2 changes as planned.

Exhibit 8. SIM-CERT logic model



SIM-CERT: Year 1 and Year 2 Changes

According to model specifications related to professional development, which were provided by the developers and the SR district implementation team, teachers in their first year of implementation were to receive two days of initial training prior to the start of the school year; and two days of ongoing training before the end of the school year. During teachers' second year of implementation, they were to receive two additional days of ongoing training within the current school year as illustrated in Exhibit 9 below. All initial and ongoing training was to be provided by SIM trainers in the first year and in collaboration with literacy coaches the second year. Initial training as planned consisted of an overview of the whole-school intervention, workshops on the Unit Organizer routine and choice of routines, as well as time for subject area teachers to collaborate on the creation of a SIM-CERT device.

Exhibit 9. Professional development by cohort: Initial and ongoing workshops as planned

	Year 1 (2006-07 school year)	Year 2 (2007-08 school year)	Total Participation in Years 1 and 2
Cohort 1	2 days (initial) 2 days (ongoing) **TOTAL = 4 days**	2 days (ongoing) **TOTAL = 2 days**	**TOTAL = 6 days**
Cohort 2		2 days (initial) 2 days (ongoing) **TOTAL = 4 days**	**TOTAL = 4 days**

Teachers also received monthly coaching from school-based literacy coaches. Monthly visits were reportedly conducted by district literacy coaches but records were not provided for review at the time of writing this report.

The Springfield-Chicopee whole-school implementation plan required the program be rolled-out in increments over the period of the grant, to achieve the goal of training a majority of teachers. Approximately 25 content-area teachers per school were to attend SIM-CERT professional development during the first and second year of implementation; 125 teachers per year. A total of 50 content area teachers per school would be trained at the end of the second year of implementation, for a total of 250 teachers.

SIM-CERT Inclusion Criteria

General efforts were to be made during the selection process to limit the exposure of READ 180 and control students to SIM-CERT trained teachers given SIM-CERT was not business-as-usual prior to the Striving Readers grant and could complicate the interpretation of impacts. Districts developed explicit criteria for selecting and prioritizing teachers for inclusion in SIM-CERT cohorts, to observe developer's SIM-CERT training requirements, and avoid potentially confounding study results.[33]

The development of criteria was complicated because developer requirements and research design considerations had to be taken into account and balanced. For example, developers initially required ELA teachers of Xtreme Reading students to be included in the SIM-CERT training, adding content to Xtreme Reading teachers' professional development. Subsequently, developers and districts determined that Xtreme Reading teachers should not receive separate training in SIM-CERT to better meet district and teacher professional development needs. In addition, some content units were not yet created for delivery. Other complications in the establishment of criteria for SIM-CERT inclusion were: (1) the same teachers delivered both Xtreme Reading and ELA in Springfield, necessitating more individual training in a very tight professional development schedule; (2) professional development in each district was both offered and conducted differently; and (3) start-up resulted in little time for explanation or buy-in for the whole-school intervention and plans.

[33] Criteria were established in consultation with evaluators. Considerations were included in the implementation and evaluation plans to ensure model fidelity would be maintained as well as the integrity of the evaluation/study within and across districts.

Therefore, criteria differed in Years 1 and 2. Districts were then to select 10th or 11th grade science, social studies, and math teachers who teach more than one class/section and do not primarily teach honors classes. If the targeted number of teachers was not met, other subject area teachers who teach 10th, 11th, and 12th grade courses were to be selected. In the second year, additional upper-grade ELA teachers were included as well as any teachers teaching 11th and 12h grade to fill training slots.

Inclusion in both SIM-CERT cohorts was to be determined within these criteria and was not planned to occur on a volunteer basis.[34] Participants were to be randomly selected from the priority groups, a more equitable process and one avoiding complications in the interpretation of outcomes given *all* teachers were eventually obligated to participate in SIM-CERT training over the period of the grant. In addition, mandatory district professional development is the reality and, therefore, is business as usual and the normal context for the SIM-CERT whole school initiative or any whole-school initiative. Teachers in the upper grades (beyond ninth grade) were to be given priority in the selection process based on the established criteria for training in both the first and second years as planned.

SIM-CERT Classroom Expectations

Finally, teachers were expected to provide explicit instruction on the routines, as mentioned previously. Teachers were also expected to co-construct routines with students to encourage and develop active learning, engagement with the subject matter, and independent mastery of the routines. In the absence of explicit guidelines and measurable expectations from SIM-CERT or districts in the first year of implementation, coaches reported developing their own expectations and tools to promote and track progress. Later in Year 1, developer guidelines indicated that, once teachers had received SIM-CERT initial training, they were expected to implement the Unit Organizer and one other Enhancement Routine for every unit.

[34] If only teachers motivated to participate were included, observed outcomes could be the result of such motivation. This selection bias is a threat to the validity of the whole school study, implemented over time. Selecting from the pool of all required participants, or those identified in groups first, is a method for avoiding selection bias and is often understood to be a more equitable way of including all teachers since all teachers were required to be trained by the conclusion of the grant.

Additionally, teachers were to integrate other Enhancement Routines as appropriate into their daily lesson plans. The different information provided by developers made for differences in training schedules between districts, and coordination among coaches made consistency across districts a challenge.

IV. Evaluation of the Implementation of the Targeted Interventions

The goals of the targeted implementation study were to present both a broad picture of the overall level of implementation and a sense of the variability that may have occurred. Differing institutional contexts or constraints influence the ways in which intervention components are implemented. Districts and schools possess their own unique complexities which may support or hinder implementation and may, in turn, affect outcomes.

The evaluation of the Springfield-Chicopee's Striving Readers Program implementation focused on …the extent to which the intensive targeted and school-wide interventions were implemented on-model," and also sought to describe the general context of implementation for the interpretation of outcomes. For this study, the extent to which an intervention is "on-model" is the extent to which the targeted intervention is implemented according to the developers' and districts' specifications and plans.[35] Implementation is evaluated within and across years.

The following sections provide the summary of the implementation study design as well as the results for this study component.

[35] Project Officer communication, November 15, 2006.

IV.A. Implementation Study Design Summary

Research Questions and Methods

Exhibits 10 and 11 include specific implementation research questions and data collection activities for the targeted interventions. The implementation research questions were developed based on the program models and their intended activities, methods, objectives, and ultimate outcome goals. Scoring is described in more detail in the following section in which implementation levels are presented.

Across the areas of implementation, data collection served multiple purposes: (1) to document and assess fidelity of implementation, (2) to determine the level of program implementation, (3) to document variation in program implementation, and (4) to examine variation in program implementation as a potential influence on observed outcomes. Data were also collected to assess the presence of relevant contextual factors for both groups of targeted intervention teachers (e.g., participation in additional professional development activities, other reform or literacy initiatives occurring at the school or within the districts). Finally, data were collected to characterize the counterfactual (i.e., what happens in the absence of a targeted intervention treatment). Although not related to the implementation of the targeted interventions, the assessment of the counterfactual—or rather what occurs as business as usual (i.e., ELA and supplemental reading supports provided)—provides contextual information to be considered in the characterization of impacts.

The multiple measures and data collection methods used for the Striving Readers' Targeted Implementation Study are displayed in Exhibit 11.

Exhibit 10. Specific implementation research questions: Targeted interventions

What was the level of implementation and variability of professional development/support for teachers/administrators?

*Professional development – **initial** training from developers:* [36]

Teachers

- What proportion of teachers received/participated at *different levels* in the initial professional development?*
- What proportion of teachers received/participated in the initial professional development at *an adequate level*?**

Administrators [37]

- What proportion of administrators received/participated at *different levels* in the professional development?*
- What proportion of administrators received/participated in the initial professional development at *an adequate level*?**

*Professional development – **ongoing** mentoring from developers:*

- What proportion of teachers received *different levels* of ongoing mentoring?*
- What proportion of teachers received *an adequate level* of ongoing mentoring?**

*Professional development – **workshops or online courses** provided by developers:*

- What proportion of teachers received *different levels* of the additional workshops or courses?*
- What proportion of teachers received *an adequate level* of the additional workshops or courses?**

What was the level of implementation and variability of classroom instruction?

- What proportion of teachers had access *to all* of the materials (i.e., technology, assessments) in time to be utilized as per the model?
- What proportion of teachers implemented the classroom model (instructional strategies/practices, schedule/pace of activities, student groupings, assessments for instruction) at *different levels* of implementation?*
- What proportion of teachers implemented the classroom model as specified by the developers at *an adequate level* of implementation?**

What characterized the counterfactual? How did the counterfactual compare to the treatment?

- What was the counterfactual (i.e., what did control students receive in the absence of treatment)?
- How did the counterfactual compare to the treatment (i.e., what students received)?

Note: In both exhibits, one asterisk (*) is used to specify cases in which components of the targeted interventions are examined by level of implementation. Two asterisks (**) are used to specify cases in which both the appropriate level of implementation and the proportion of teachers evidencing this level of implementation were used to examine intervention implementation.[38]

[36] Initial training for teachers is defined as training that took place in the planned summer professional development (PD) period prior to the second year of implementation. This PD is considered to be the foundation for program implementation. Administrators received initial PD at the beginning of the school year in Year 1. There was no formal training for administrators in Year 2 for either targeted intervention.

[37] Facilitators (district staff members) were trained as per the READ 180 model in the initial PD sessions to support the Scholastic RED online courses in Year 1. In Year 2, there was no such training as the remaining facilitator (only one facilitator in Year 2) had been trained in Year 1.

[38] Exhibits were developed by Abt Associates, the technical assistance provider to Striving Readers evaluators.

Exhibit 11. Research questions and data sources: Targeted implementation study

Research Questions*	Measures/Data Sources**					
	Surveys/ Interviews		Observations	District Records/ Records Review		
	Teacher	District-school administrative staff	Teacher (classroom)	Professional development attendance	Curricula, district-provided developer materials	Class rosters (scheduling, dosage)
What was the level of implementation and variability of professional development/support for teachers/administrators?						
Professional development/support (PD) for teachers (initial, ongoing mentoring, and workshops and courses)						
INITIAL PROFESSIONAL DEVELOPMENT						
Proportion of teachers receiving different **levels** of *initial* professional development*	√			√	√	
Proportion of teachers receiving **an adequate level** of *initial* professional development**	√			√	√	
ONGOING PROFESSIONAL DEVELOPMENT – MENTORING						
Proportion of teachers receiving different **levels** of *ongoing* professional development via mentoring*	√			√	√	
Proportion of teachers receiving **an adequate level** of *ongoing* professional development via mentoring**	√			√	√	

Research Questions*	Measures/Data Sources**					
	Surveys/ Interviews		Observations	District Records/ Records Review		
	Teacher	District-school administrative staff	Teacher (classroom)	Professional development attendance	Curricula, district-provided developer materials	Class rosters (scheduling, dosage)
ONGOING PROFESSIONAL DEVELOPMENT - WORKSHOPS, SEMINARS AND COURSES						
Proportion of teachers receiving different **levels** of professional development *workshops/courses**	√			√	√	
Proportion of teachers receiving **an adequate level** of *workshops/courses***	√			√	√	
What was the level of implementation and variability of classroom instruction?						
Proportion of teachers with access to all materials (e.g., technology, assessments)*	√		√			
Proportion of teachers who implemented the classroom model at different **levels***	√		√			
Proportion of teachers who implemented the classroom model at **an adequate level****	√		√			

Targeted Implementation Data Collected

The implementation data collected via each method described in the previous table is detailed below. Measures used are provided in the appendices. Evaluators collected primary data twice per year based on the schedule established in the initial year. District agreements were made with teaching staff (supported by Striving Readers funds) to provide the necessary evaluation data. In addition, districts required other staff with knowledge of Striving Readers implementation or knowledge of the "counterfactual" to participate in data collection activities. The SR district implementation team supported evaluator efforts to obtain complete data as well as provided secondary data they collected while documenting implementation activities.

Teacher Surveys

All targeted intervention teachers (ninth grade and upper grades) were asked to complete the Striving Readers survey via an Internet-based survey provider in Year 1 (May 2007) and in Year 2 (May 2008). An initial e-mail was sent (to addresses provided by the SR district implementation team) notifying targeted teachers of the upcoming survey. A subsequent e-mail, including the embedded link to the survey teachers were to complete, was sent, and two additional follow-up emails were sent to increase response rates. In Year 1, the response rates for the Xtreme Reading and READ 180 teacher surveys were 70% (7 of 10) and 91% (10 of 11), respectively. In Year 2, the response rates for Xtreme Reading and READ 180 surveys were 100% (12 of 12) and 100% (13 of 13), respectively. The higher response rates in Year 2 were likely the result of more familiarity with the process and the fact that no teachers were absent long-term, as had occurred in the initial year.

The purpose of the READ 180 and Xtreme Reading teacher surveys was to assess teacher-reported: (1)adequacy of materials (teacher and student) provided to implement the intervention (2) level of compliance in the implementation of intervention components within their classes (e.g., sequencing of curriculum, instructional strategies, frequency of assessments, frequency of adaptations/changes to the intervention); (3) professional development participation and satisfaction with training offered; and (4) barriers or challenges affecting their classroom-level implementation of the interventions (surveys are included in Appendix B).

Teacher and Administrator Interviews

Interviews were conducted with target intervention teachers, control classroom teachers, school and district administrators, as well as other key staff in May 2007 (Year 1) and May 2008 (Year 2). In Year 1, all intervention teachers were interviewed (both ninth and upper grades) but only ninth grade teachers were interviewed in Year 2. Additional interviews were added with different groups of stakeholders in Year 2 in order to further characterize implementation. As a result, a tradeoff was made to conduct teacher interviews in the second year only with those involved in the randomized control trial (RCT)—ninth grade teachers.

Interviews (see Appendix C) were conducted in-person with READ 180 teachers and Xtreme Reading teachers to gather information which did not lend itself to incorporation in the online survey or could not be obtained via classroom observations. The READ 180 and Xtreme Reading targeted intervention teacher interviews were used to learn more detail about: (1) teacher-reported implementation successes and challenges, (2) the *nature* of the intervention adaptations made, (3) factors affecting pacing (how quickly teachers were able to move through the program), and (4) how school and district policies or programs affected implementation.

Individual interviews were also conducted with ninth grade control classroom teachers in order to learn about the ELA ninth grade courses offered at each of the five high schools to all targeted study students.

More specifically, the purpose of the interviews was to characterize general curriculum and instruction[39] and supplemental reading supports (i.e., establish a treatment contrast or counterfactual). Additionally, the interviews included items asking about in-class and out-of-class supports for students having difficulty in reading prior to the grant and during each year of implementation. The interviews concluded with questions about contamination (e.g., whether or not the teacher received training in either of the two interventions, uses intervention practices and/or materials or assessments).

Interviewers asked key school and district administrators, ELA chairs and supervisors, guidance counselors, special education directors/supervisors, and Instructional Leadership Specialists, about their roles and responsibilities as they related to the Striving Readers program (refer to Appendix E). The interviewer included questions about implementation successes and challenges, district and developer roles and support, district and state policies influencing implementation, school and district context (e.g., previous literacy programs or reform efforts), and perceptions of teacher implementation and student and teacher outcomes.

Interviews included items for ELA chairs (school staff), ELA supervisors (district staff), guidance counselors, and special education directors to characterize business as usual (i.e., the control classroom ELA experience and any additional supports). The interviews were conducted to establish what, if any, supplemental services were offered or experienced by struggling readers in the absence of the treatment (i.e., the counterfactual). Guidance counselors were asked about the process of scheduling students assigned to treatment or control interventions; their role in the scheduling, placement procedures, and verification; and their recommendations regarding key staff members to speak with for more information regarding placement. Special education directors, supervisors, and staff were asked to describe: (1) what programs or classes were generally offered to students in need of extra reading instruction, (2) the students who were receiving supplemental services, and (3) the students ineligible for or excluded from the Striving Readers program.

[39] Items related to curriculum and instruction were constructed to obtain information about core components of the English course (i.e., reading and writing), lesson plan development, grading and student assessment, grouping of students for instruction, and approaches to teaching writing and reading.

Exhibit 12 displays the number of interviews conducted in both Year 1 and Year 2. The numbers interviewed of administrators and faculty involved in the Striving Readers project provided a comprehensive picture of the implementation in the first and second year of the project. In Year 1, interviews were planned for 48 staff members and 100% of these interviews were completed (56 total staff members were involved in Striving Readers; 85% were interviewed). In Year 2, interviews were also planned and conducted with 48 staff members but the population had increased (68 total staff members were involved in Striving Readers; 71% were interviewed).

Exhibit 12. Number of interviews by stakeholder group (May 2007 and May 2008)

	Year 1		Year 2	
	Number interviewed May 2007	**Number in Districts SR**	**Number interviewed May 2008**	**Number in Districts SR**
Xtreme Reading teachers	7	10	5*	12
READ 180 teachers	13	13**	5*	13
Control classroom teachers	4	5	5	5
Principals or Vice-Principals	5	5	5	5
Guidance counselors/schedulers	5	5	4	5
ELA department chairs	5	5	4	5
ELA directors/curriculum directors	2	2	2	2
Reading Instructional Leadership Specialists	--	--	2	3
Assistant superintendents/superintendents	2	4	2	4
CERT literacy coaches	5	5	5	5
Special education supervisors (school)	--	--	5	5
Special education directors (district)	--	--	2	2
SIM and READ 180 model developers	--	--	2	2
Total	**48**	**56**	**48**	**68**

*In Year 2, only 9th grade RCT teachers were interviewed, whereas all teachers were interviewed in Year 1 (including the upper grade teachers).

**Note: In Year 1, one READ 180 teacher was promoted to an in-district position mid-year and was substituted with another READ 180 trained teacher. So although there were a total of 12 teachers at any one point in time, the total number of teachers interviewed was 13.

While the number of interviews conducted remained constant from Year 1 to Year 2, the pool of potential interviewees was broadened in the second year: Special education staff were added to obtain more information about the counterfactual and the Reading Instructional Leadership Specialists (ILS) were added to provide more information about classroom level practices. In the fall of 2008, evaluators also conducted interviews with Xtreme Reading and READ 180 model developers to obtain information about implementation in general and about the changes made to the intervention model (i.e., changes to materials, pacing of instruction, professional development content, etc.) from Year 1 to Year 2.

Classroom Observations

In Year 1 and Year 2, teacher classroom observations (refer to Appendix D) were conducted by evaluators twice during the school year—in February and May 2007, and in February and May 2008—in order to collect data on classroom-level implementation.[40] In Year 1, classroom observations of Xtreme Reading and READ 180 teachers (9th grade RCT and upper grade classes) and control classroom teachers were conducted. In Year 1, a total of 25 individual teachers were observed and a total of 40 observations were conducted for those teachers during the course of the year. In Year 2, classroom observations of Xtreme Reading and READ 180 teachers (9th grade RCT classes only) and control teachers were conducted. A total of 15 individual RCT-only teachers were observed during the 2007-08 school year and a total of 28 classroom observations were conducted for those teachers.

District Records (Professional Development Attendance and Other Materials)

Additionally, secondary data and extant documents provided by districts to document their implementation efforts were collected for analysis by evaluators.

[40] The original schedule for data collection as planned was fall and spring. However, the initial period was determined by the dissemination of the grant and time required for the startup of implementation, as well as increased emphasis on the implementation study, which necessitated the development of tools.

For example, the districts provided records of professional development attendance (initial professional development, ongoing professional development and in-class coaching participation) for both READ 180 and Xtreme Reading. For READ 180, district staff provided a computer-generated progress report of teacher participation in the online sessions for the RED Course.[41]

Miscellaneous documents reviewed include developer materials, meeting minutes, memoranda, written curricula, and course syllabi. In addition, many meetings were held with the districts as well as clarifications made post-meetings and receipt of data. Documents were used to corroborate findings and for triangulation purposes.

Targeted Implementation Teachers

Although districts decided they must hire teachers as individual district employees, they agreed to use the same job description to ensure that any teacher qualified for hire would be considered qualified *across* districts. The job description per the Implementation Plan listed the qualifications for new teacher hires, including: (1) certification in English or reading or in the process of attaining either; (2) five years of experience in teaching English or reading; and (3) some experience in use of technology and availability to attend summer professional development training. In addition, teachers applying for the positions and those hired were to agree to be randomly assigned to one of three conditions: Control, READ 180, or Xtreme. Teachers could not request or select what they were to teach as per hiring requirements. However, these agreements were not explicitly included in the teacher contracts; standard district teaching contracts were used.

[41] For the Year 3 report, evaluators hope to obtain the district records regarding READ 180 implementation via the Scholastic Achievement Manager and their on-site servers.

In Year 1, there were six READ 180 teachers, five Xtreme Reading teachers, and five control classroom teachers for a total of 16 included in the RCT. In Year 2 there were five READ 180, five Xtreme Reading, and five control classroom teachers for a total of 15.[42]

Characteristics of READ 180 Teachers: Year 1 and Year 2

In Year 1, none of the READ 180 teachers reported experience teaching this intervention prior to their employment now supported by this grant. When surveyed about their licensure, (five of the six ninth grade teachers responded to the survey), three teachers reported having a professional license and one reported having a provisional license.[43] The remaining teacher responded she/he had an "initial" license. Years of teaching experience for the five respondent teachers ranged from one to twelve years, with two teachers reporting four years of experience, and one teacher reporting seven years. Four of the six READ 180 RCT teachers in Year 1 had five or more years of teaching experience.[44] The highest teaching degree completed as per the teacher resumes was an M.A. degree in the field of education (four teachers). The remaining teachers did not obtain graduate degrees but did indicate having a B.A. in English (two teachers).

In Year 2, one Year 1 teacher remained a ninth grade READ 180 teacher. However, two of the five teachers newly included in Year 2 had experience teaching READ 180 because they had taught READ 180 to upper-grade non-RCT students in Year 1. None of the teachers reported having had previous experience teaching READ 180 prior to their grant-funded positions. When surveyed about their licensure, all five of the Year 2 teachers responded that they held initial certification. In Year 2, there were two teachers new to the field of teaching—having taught one or two years in all. The other three teachers had more than five years of teaching experience (more specifically, teachers reported six, eight, and twelve years of experience).

[42] As described, teachers delivered interventions in upper-grades (10th and above) but there was not a control group included in these grades as per district plans (so these were non-RCT teachers). There were reportedly 30 Striving Readers teachers in Year 2: 12 READ 180, 13 Xtreme Reading, and 5 Control. In Year one, the total including upper grades was 12 READ 180, 10 Xtreme and 5 Control.

[43] Although teacher resumes were provided, the way in which licensure and qualifications were reported differed so items were included in the survey to capture this information in a more systematic way for analysis and reporting purposes.

[44] Data on type of previous experience, e.g. English or Reading were not available for all teachers.

The highest teaching degree completed as per teacher resumes was an M.A. degree in Year 2 as well, but fewer teachers reported having this degree (two teachers, one in Education and one in Secondary Teacher Education). The remaining teachers reported completing a B.A. (two in English and one in Education).

All Year 2 teachers were relatively new to the school—when surveyed in the spring of 2008, three teachers responded that they worked at the school for two years and two teachers stated it was their first year at the school. When asked about the number of READ 180 sections taught, teacher responses ranged from one to two sections of READ 180 in Year 1 and ranged from one to three sections of READ 180 in Year 2.

In Year 2, *all* teachers held initial licenses and while two were new teachers, three met the "five or more years of experience" criteria set by the districts. In Year 1, half of the teachers, three of the six, had a professional license and four of the six met the "five or more years of experience" criteria set by the district.

Characteristics of Xtreme Reading Teachers: Year 1 and Year 2

In Year 1, Xtreme Reading teachers did not report having had experience teaching Xtreme Reading prior to the grant. When surveyed about their licensure (three of the five ninth grade teachers responded to the survey), three reported having a professional license. There are no data in this regard for the remaining two teachers who did not respond to the survey. Years of teaching experience for the five teachers (per survey and resumes) ranged from two to twenty years, and three of the five teachers had five or more years of teaching experience. In Year 1, the highest teaching degree completed as per teacher resumes was an M.A. degree (one in Education and one in Reading). The remaining three teachers did not obtain graduate degrees, but did report having B.A. degrees in English, Liberal Arts, and Law respectively.

In Year 2, two Year 1 teachers remained ninth grade Xtreme Reading teachers. An additional teacher newly included in Year 2 had experience teaching Xtreme Reading because she/he had taught Xtreme Reading to upper-grade non-RCT students in Year 1. So, in Year 2 of the Xtreme Reading intervention, three of the five teachers had experience teaching Xtreme Reading (two taught ninth grade students in Year 1 as part of the impact study while one taught the upper grades that same year). None of the teachers reported having had previous experience teaching Xtreme Reading prior to their grant-funded positions. The remaining two of the five Xtreme Reading teachers were new in Year 2 of the study.

When surveyed about their licensure, a range of certification levels was reported: two teachers reported "initial," one "provisional," one "professional," and the remaining "preliminary."[45] Three teachers have a B.A. in English (one of which has a Ph.D. in Jurisprudence). The fourth teacher has a B.A. in Law and the fifth has an M.A. in Education. Three of the teachers were new or relatively new to the field of teaching (having taught one, two, and three years). The other two teachers had seven and fifteen years of experience, respectively. Two teachers were completely new to the school in 2007-08, while two had been there for two and three years respectively. One teacher had worked at the school for a total nine years.

As reported, the number of Xtreme Reading sections taught ranged from one to five. In Year 2 the number of Xtreme Reading sections taught by each teacher ranged from two to four sections. In Year 1, more than half of the teachers, three of the five, held a professional license at a minimum; three of the five met the "five or more years of experience" criteria set by the district. In Year 2, only one teacher held a professional license and two teachers had taught for five years or more.

[45] Preliminary licenses are for people who have not completed Approved Educator Preparation Program but who hold a Bachelor's Degree, have passing scores on MTEL, and have completed additional course work for some (?) licenses. Initial licenses require a Bachelor's Degree, passing scores on MTEL, and the completion of an Educator Preparation Program. The professional license requires three years of employment under an Initial License, Completion of a Teacher Induction Program and options for the Professional License (downloaded from MA Department of Elementary and Secondary Education, March 2009).

Characteristics of Control Teachers: Year 1 and Year 2

Four of the five control classroom teachers in Year 1 continued as control classroom teachers in Year 2, the highest number of returning teachers of the three groups (Control, READ 180, and Xtreme). Therefore, one teacher was new to the control group in the second year of the study. In terms of overall teaching experience (per resumes at the time of hire), years of experience ranged from no experience to eight years of teaching experience—with three teachers reporting two, three, and four years of experience.

Per the district, all of the Control teachers in Years 1 and 2 were considered to be "highly qualified" meaning that they have: (1) a Bachelor's degree; (2) a teacher's license (regardless of type); and (3) proof that they have mastered their content as evidenced by passing scores on the Massachusetts Tests for Educator Licensure (MTEL's), for example. The highest teaching degree completed as per teacher resumes was an M.A. degree in English (one teacher). Four of the five control classroom teachers have a B.A. in English.

The number of English Language Art (e.g., creative writing, publications production, etc.) classes taught by each teacher in Year 2 ranged between three and five, depending on the district in which they work.

Business as Usual

The counterfactual is addressed by the inclusion of a control group to answer the question, "What would happen in the absence of treatment?" In the case of the targeted study, the control group was a business-as-usual group receiving whatever supplemental reading supports that would normally be provided. The two components of business as usual for the Striving Readers study include (1) the standard ELA courses all students receive, and (2) the supplemental services ordinarily available to students in need of additional reading support.[46]

[46] Note that business as usual globally consists of all course requirements for graduation as well as exposure to school and district-wide initiatives. Only those courses and initiatives implemented specifically to enhance literacy were described given the purpose of this initiative.

The second component was the true counterfactual because, as per final cross-district plans to ensure consistency, the treatment was an add-on or supplemental service, provided in addition to required ELA courses; all students, treatment included, were to receive the standard ELA course.[47]

An analysis of data collected from district documents, interviews of control classroom teachers and administrative staff, and observations of their classes allowed evaluators to note how course content was planned and delivered; what instructional strategies were employed by control teachers; and which instructional supports were provided to struggling readers during, and in addition to, the standard ELA class period. Finally, these data were used to determine any potential study "contamination" (i.e., the incorporation of targeted intervention materials in class or reported training experiences similar to those of targeted intervention teachers).

English Language Arts

In Chicopee, curricular goals for students are based on state standards in Language, Reading and Literature, and Composition and Media Literacy. Goals are also described in the district scope and sequence documents and in documents listing curricular benchmarks. Based on a review of scope and sequence documents, students in Chicopee read more in-class selections, though both districts shared some common texts (e.g., *Of Mice and Men*, *To Kill a Mockingbird,* and *Romeo and Juliet*). Chicopee's ninth grade curriculum includes lengthier selections in general (e.g., two Greek plays). There is an emphasis on grammar instruction and specific goals for vocabulary instruction, including teaching students to use word roots to derive contextual meanings. Chicopee reportedly updates benchmarks by term each year, though the defined content has remained the same since 2004.[48]

[47] Students identified as struggling readers included Students with Disabilities (SWDs) and English Language Learners (ELLs).

[48] Assessments are described in the broad language of the state standards. The curricular "benchmarks" for each term and strand are located in a separate document and updated annually. This document outlines single assessments or points of emphasis for each unit and marking period, allowing teacher flexibility in choosing when to address the curricular goals. Some specific district-wide assessments are described such as business letter writing or expository essays, though the content is left to the teacher to determine.

Chicopee teacher syllabi used for standard ELA courses were reviewed and remained largely unchanged from Year 1 to Year 2 and followed the same guidelines for use of district-provided materials.

The ways in which the curriculum is delivered can vary from teacher to teacher given the district does not provide specific daily lesson plans and does not prescribe specific classroom activities. Interviews with Chicopee administrators indicate that while there may be variation in how district curricular materials are applied in the classroom, the accreditation process dictates that curricula across the two high schools be parallel. As one administrator stated:

> *The syllabi of the schools might be different; they might use a different resource, or a different textbook, or a different trade book to teach a skill or strategy, but it's all aligned to the curriculum and the genre. If they're studying, say, drama – these are the skills that children should learn, and then these are the resources and tools that the schools will have to use to do so.*

In Springfield, the second year of implementation oversaw changes in the degree of rigor and standardization of the ninth grade ELA curriculum. According to ELA department chair interviews, several additions of text were made to the ninth grade curriculum, including a nonfiction text and play teachers could choose from predetermined selections. An internal document highlights specific ELA strands, guiding principles, and learning standards, as well as instructional strategies for block period scheduling and a pacing guide for major units of study for each grade level.[49] The text selections and dates for the coverage of selections are detailed, as are suggested lesson plans. The document or guide provides daily lesson plans for each of the five major units matched with learning standards. This document also includes suggestions about lesson design including a "block activator," which functions as a theme or essential question for the class period, a mini lesson, learning activity, wrap up, and homework.

[49] The pacing guide lists required and voluntary reading selections, identifies standards that should be addressed during these units, and provides broad descriptions of assessments that should accompany each unit.

The suggested lessons include strategies for differentiating instruction such as mini lessons and homework activities focused on vocabulary development, or pre-reading activities intended to help students understand thematic or historical context.

Overall, they are paced for quick coverage and rely heavily on students doing work outside of class time. In fact, almost all reading is assumed to take place outside of class (teacher interview data corroborate time is not allotted for in-class reading and teachers also noted that students often come to class unprepared).

Teacher interview data indicated wide variability in how district plans were implemented among Springfield schools. At one district school, a teacher explained, *"…people in our building aren't necessarily following the district lesson plans to a T; they're using their own modifications or…adaptations."* The perception at another school was that teachers were expected to follow the prescribed lesson plans with fidelity: *"we're supposed to exactly teach the lesson plan as it's given."*

Control Observations

Control classroom observations illustrate more consistency across districts and schools in the instructional focus and materials employed than one might expect, given the noted curricular differences or changes based on the review of district curricular materials as well as interview comments.[50] All lessons included either some discussion of new vocabulary or recounting of plot in addition to the discussion of basic plot. In-class writing was infrequently used in the observed classes, though the writing assignments that were mentioned included summary, autobiographical, and expository writing. Also consistent among control classrooms within the two districts was the pedagogical approach to teaching the ELA curriculum; that is, the instructional strategies teachers use to deliver the course content.

[50] While observed lessons in February 2008 involved a variety of texts, including *Of Mice and Men*, *West Side Story*, *Black Boy* and a mythology unit, observed classes in May were all using *Romeo and Juliet*.

Classroom observations conducted in February and May in Year 2 show that full-class reading and discussion was the most commonly used teaching method, though pair and individual work was also occasionally assigned. Most observed lessons were teacher-centered, with the teacher as the reader, lecturer, or discussion leader for the majority of the period, with occasional breaks for individual vocabulary work or other seatwork. Reading skills such as decoding and summarizing were assumed.

Among the instructional strategies used to cover curricular material, teachers primarily emphasized vocabulary and comprehension skills. New vocabulary or literary terms were defined and presented in every observed lesson, in formal or informal ways. Teachers used a wide variety of vocabulary enrichment strategies, including direct instruction on new vocabulary, practice for upcoming vocabulary tests, and explicit definitions of literary terms. Teachers also expected students to summarize reading passages in their own words to increase their comprehension of assigned texts. In both sets of observations, most teachers relied on closed questioning to elicit responses about reading passages from students in a full-class format, as both an informal assessment of students' understanding of the literary work and a means of emphasizing important factual points about plot or characterization. Finally, teachers in control classrooms were not observed to be providing reading instruction tailored to struggling students; most classroom activities involved reading and responding to a work of literature.

Class Organization, Structure, and Context

As reported in the Year 1 executive summary, there were planned differences in the dosage of standard ELA based on several factors: each district's scheduling; supplemental reading supports inclusive or exclusive to ELA instructional time; and graduation and course requirements. These differences were outlined in the districts' implementation reports and illustrated in the final dosage chart (see Exhibit 13). There were also unplanned differences noted in ELA dosage in Springfield where the total length of the ELA course per day was not delivered as anticipated, resulting in differences in the total average dose delivered.

Exhibit 13. Business-as-usual ELA for Striving Readers across districts (Years 1 and 2)

	CHS	CCHS	HST	COMM	PUTNAM A Week	PUTNAM B Week
PLANNED IMPLEMENTATION						
Control Group: Business as Usual*	45 min ELA + whatever additional support normally provided	45 min ELA + whatever additional support normally provided	90 min ELA block inclusive of support normally provided	90 min ELA block inclusive of support normally provided	90 min ELA block inclusive of support normally provided	
ACTUAL IMPLEMENTATION						
Control Group: Business as Usual*	45 min ELA + whatever additional support normally provided	45 min ELA + whatever additional support normally provided	90 min ELA block inclusive of support normally provided ***(1st or 2nd Semester but not both)***	90 min ELA block inclusive of support normally provided ***(every other day for year)***	90 min ELA block inclusive of support normally provided	***45 min ELA*** block inclusive of support normally provided

Note: Business-as-usual ELA was to be provided to all Striving Readers: READ 180, Xtreme Reading, and Control. The information provided for this chart as reported in 2007 remained the same for Year 2.

Observed differences that related to school schedule restrictions were district-specific and included: (1) a vocational-technical school operating on A and B weeks and (2) two schools operating on differing block schedules in which ELA was provided either every other day or in only one of two semesters. Refer to the overall summary of the targeted implementation study below for more details regarding these challenges. Year 2 observation data confirm school schedule records and district variation in dosage as reported in Year 1. In Chicopee, observed classes ranged from 45 to 52 minutes in length. The time afforded to Springfield control classes varied as per the exhibit. Two of the observed classes were scheduled as portions of 90 minute block periods. The control class at the vocational school operated on an A/B week schedule with observed class times ranging from 40 to 90 minutes.

In addition to the dosage or scheduled time for standard ELA classes, teacher-student ratios were also considered to be a structural element that characterizes the business-as-usual service that students received.

Control classroom teacher interviews and observations conducted by the evaluators in February and May of both 2007 and 2008 point to a great deal of variation in control classroom size. In Chicopee, observed attendance in control classes ranged from 18 to 27 students (class sizes were to be capped by districts, e.g., 25 students), but in Springfield, fewer students were observed to be present. Three of the observed control classes had four, five, and nine students, respectively. The largest control class observed had 18 students in the vocational training school.

Counterfactual (Supplemental Services)

The counterfactual consists of normally provided services that struggling readers, in the absence of Striving Readers, received above and beyond standard ELA and reading instruction. In Year 1, none of the five high schools participating in the Striving Readers study had a comprehensive approach to support students in need or struggling readers other than the school-wide implementation of SIM-CERT afforded by this grant. As in Year 1, in Year 2 there was no comprehensive approach in the identification and delivery of supplemental supports for these students, and the existing interventions that were provided were often limited.

Refer to the exhibit below for a summary of the supplemental services normally provided to struggling readers in districts.

Exhibit 14. Supplemental reading support available to struggling readers (2006-08)

	Description of Striving Reader Population	Description of Support Offered	Chicopee Course Title	Springfield Course Title
General Striving Reader	Students identified/referred based on prior achievement test scores (including MCAS); teacher or guidance counselor referral; or student request	Supplemental reading and writing instruction (small group and individual) MCAS preparation - tutoring Targeted intervention program (minimally implemented – only select schools)	*Individualized Reading and Writing* or *Reading Improvement* *MCAS English Review*	*Reading and Writing Lab* - optional and offered to all (in only one school) *READ 180 v1.6 non-Enterprise* (in only one school)
ELL Striving Reader	Limited English Proficiency (LEP) students enrolled in the *English For Speakers of Other Languages (ESOL)* courses [51]	English language learner support in addition to instruction in ESOL course	*ELL Learning Skills* (in only one school)	Not offered in 2006-07 school year *Reading and Writing English* course added in 2007-2008
Special Education (SPED) Striving Reader	Students with Individualized Education Program (IEP) requirements – or team determines need based (not all SPED students required to take these courses) [52]	Targeted intervention programs including summer programs (only select schools) Study skills courses – MCAS preparation (only select schools)	*READ 180 v1.6 non-Enterprise* *Lindamood Bell*	*READ 180 v1.6 non-Enterprise* (in only one school) *Study/Learning Skills/Reading Improvement* -restructured as the *Reading and Writing Enrichment* course for 2007-08 (in two schools) *Summer Reading Clinic*

Most schools provided some form of state test (MCAS) preparation; students were identified for this supplement by school or district-wide assessments or by their MCAS scores. Otherwise, students were responsible for seeking out tutoring with their teachers or attending occasional after school tutoring sessions.

[51] In SPS "students are assigned to a Sheltered English Immersion instructional setting based on their language dominance as well as their academic proficiency." Designation as an LEP student and placement in ESOL is based on the parent report of home-languages, oral language proficiency level using the Bilingual Syntax Measure (BSM), as well as the level of literacy based on a reading inventory. Refer to http://sps.springfield.ma.us/deptsites/ell/Questions.asp.

[52] Team is defined as the Education Team Leader (ETL) for the school, the general education teacher, the SPED teacher, the school psychologist, any specialists such as speech or occupational therapists, and the student's parent. Placement may also be based on IQ and WIAT-II assessment results.

In Year 2, general struggling readers who do not have diagnosed special needs reportedly received few supplemental services for reading outside of standard ELA classes in both Chicopee and Springfield. In general, only those classified as Special Education students or English Language Learners (ELLs) had access to additional literacy support.

According to interviews with teachers and district personnel, supports for general struggling readers were unchanged from Year 1 to Year 2.[53] One administrator in Chicopee noted that the school's involvement in Striving Readers had increased awareness about the problems of poor reading among staff and teachers.

SPED Designated Students and English Language Learners (ELLs)

Year 2 data show that some students received reading support as a result of special education services outlined in individualized education plans (IEPs) or based on Limited English Proficient (LEP) status, though the specific interventions varied from school to school. Lindamood Bell was utilized in one district and READ 180 v1.6 was used in both for a limited number of Special Education students. Although one district reported using READ 180 for six to eight years within the Special Education (SPED) department, not all components of the program were utilized. In the 2007-08 school year, the content of study skills courses was changed to serve as general enrichment courses required for SPED designated students identified by "prior standardized testing" as being at-risk of failing the state assessment. These courses were required for all students failing the state assessment administered 10th grade. District staff reported that teachers used MCAS item analysis in these courses to identify student strengths and weaknesses as readers and writers, and to tailor ELA instruction based on student text responses. In one district, a summer reading clinic was offered to SPED designated students receiving services during the school year to reduce the risk of substantial regression over the summer in the absence of those services.

[53] One vice principal reported a plan to replace the existing reading and writing program with PLATO, computer assisted software intended to develop reading skills, during Year 3 of the Striving Readers grant (2008-2009 school year).

ELL students identified in need were required to enroll in the English for Speakers of Other Languages (ESOL) courses in addition to other requirements. Two of the high schools added these courses in the 2006-07 school year. The courses provided tutoring and additional support to ELLs in the development of their reading and writing skills, receiving content similar to the content targeted students received.[54]

Contamination

In general, neither Chicopee nor Springfield control teachers reported receiving training relating to reading development or instructional strategies for supporting struggling readers in Year 1 or Year 2, consistent with reports in Year 1.[55] As one control classroom teacher explained, *"There's a lot of PD [professional development] there but we haven't really had anything on reading…I've never really had any training in teaching reading, ever."*

In both districts, there was no evidence of the contamination of control classrooms, defined as an infusion of targeted materials or instructional strategies in Years 1 or 2. Control classroom teachers were not observed to be using the current READ 180 and Xtreme Reading materials, technology, or model-specific instructional strategies in any of the classrooms, nor did they report using them. The unique characteristics of the interventions were not found to be incorporated in the supplemental services control students received.

As described previously, instruction with an earlier version of READ 180, version 1.6, was provided in one district to a small percentage of special education students as per their individualized education plans (these were the business-as-usual supplemental services for these students). In addition, prior to entering high school, the other district provided a small percentage of students, primarily special education students, with these services as well (approximately 15% as reported by the district). However, according to the districts there was no differential receipt of services for those students later among the three groups: READ 180, Xtreme Reading, and Control.

[54] Interviewed teachers indicated some students reportedly received help from Saturday school sessions but the number was believed to be very limited (this was not an option offered in district documentation but there was mention of additional school time for particular circumstances related to attendance or behavior.)

[55] Two of the three Springfield control classroom teachers reported attending training in the John Collins Writing program.

Teachers assigned to teach standard ELA to classes of control students in Year 2 either stated that their students had no other reading support outside of what was provided in regular ELA classes (unless they were identified as special needs students) or that student participation in the occasional tutoring opportunities at the school was voluntary.

As shown in Exhibit 15, the potential to use READ 180 v1.6 for regular-education populations and to implement in regular-education classrooms did not occur.[56] This exhibit provides a breakdown of the existing packaged interventions provided by school, all primarily provided to students with special needs.

Exhibit 15. Existing literacy interventions (2006-07)[57]

School	Initiative	Length of Implementation	# Served	Frequency	Duration	Degree of Success
Chicopee High	READ 180	1 year	20	Daily	90 min	Sporadic
	Lindamood Bell	1 year	16	Daily	48 min	Minimal
Chicopee Comp	READ 180	3 years	42	Daily	90 min	Unknown
	Lindamood Bell	1 year	32	Daily	48 min	Unknown
Putnam	READ 180	5 years	102	Daily or bi-weekly	48 min	Min. of a grade level per yr.
SciTech	READ 180	2 years	120	Daily	70 min	Sporadic
Commerce	READ 180	3 years	100	Weekly	90 min	Unknown

Year 2 data provided more information on the types of supplemental services available to students during the first two years of implementation.

Business as Usual Summary

In summary, there was little change in the ELA curriculum from Year 1 to Year 2 in Chicopee. There were substantial changes in the ELA curriculum from Year 1 to 2 in Springfield because the district sought to increase consistency across schools in terms of the curriculum (reading and writing requirements), pacing, and use of standards.

[56] Estimates provided in this exhibit were based on district reports of potential use.
[57] A version of this exhibit was originally included on page 13 of the Springfield-Chicopee Striving Readers Program Implementation Plan (submitted June 2006).

Additionally, while supports were provided to select struggling readers in a variety of forms across the districts, there continues to be no systematic district-wide approach to identifying and delivering supports to Striving Readers. In general, students classified as Special Education students or ELLs had the most access to additional literacy support outside of standard ELA classes. In the absence of such designation, however, the availability of supplemental supports for students was minimal.

IV.B. Level of Implementation

Implementation levels characterize implementation context and complexity in a meaningful, clear and understandable way.[58] In addition, defining levels of implementation provides a way to gauge the magnitude of an identified influence on study outcomes. Therefore, a systematic approach was used to define measurable and meaningful facets of the interventions to be rated according to proposed specifications for implementing the Striving Readers Program.

Ratings serve the purpose of providing a snapshot of implementation level rather than an accounting of every nuance of implementation.[59] Implementation scoring is a descriptive process and is not intended to predict (or directly connect to) the impact of the interventions, which are being studied precisely because those impacts under the described conditions are unknown. In addition, data were collected in snapshots and by definition represent only a picture at that point in time. Finally, it is important to note that the interventions were not equivalent and therefore their ratings should not be compared.

[58] It is important for implementation plans to be clearly defined to allow for a systematic assessment of implementation with the goal of gaining an understanding of ways in which context may influence outcomes.
[59] These nuances, though difficult to measure or document, represent potentially important aspects of the interventions.

Methods

The development of a framework for describing and rating implementation began early in the first year of implementation with the development of intervention logic models. The models were based on developers' specifications for the interventions and other information obtained from developers and districts.[60] Ratings were derived from data obtained via fidelity of implementation measures, which included measures provided by intervention developers and modified by evaluators for research purposes. Teacher observation data were used whenever possible to assign ratings and survey data were used when observation data were not available or sufficient. In addition, the districts provided data necessary to calculate the ratings for the delivery and receipt of professional development.

The process of identifying a framework for the levels of implementation was challenging given the overall complexity of the interventions. The development and evaluation of the level of implementation involved three major phases. The first phase required the identification of the key measurable components of the targeted interventions. The second phase involved the specification of measurable subcomponents and indicators for each component and the rating of each subcomponent based on developer-model and district plans. The third and final phase involved the calculation of an aggregate score for the components in order to determine the overall adequacy of implementation of the intervention model.

Phase 1: Implementation Components

Intervention logic models provided the necessary framework for identifying the key components of the targeted interventions to be assessed for implementation fidelity.

[60] Although ratings were to be developed as part of the implementation study and reported in the second year, it was later required as a result of expanding the depth and scope of the implementation study that ratings be reported in the first year as well.

The logic models reflect what was planned" by the districts in conjunction with the model developers and thus what was "required" for adequate implementation. (Note that the terms planned and required are used interchangeably in this report.)

As per the logic models, each intervention encompasses both specifications related to classroom model implementation (e.g., curriculum covered during class time, use of instructional practices, use of assessments to inform instruction) and specifications related to the necessary inputs for achieving classroom implementation (e.g., professional development training for teachers, materials, scheduling of time required for implementing the intervention, teacher-student ratio).

Based on these specifications, the following five components were identified to assess the fidelity of implementation of the targeted interventions.

1. Professional development (inputs)[61]
2. Materials, technology, assessments (inputs)
3. Classroom organization, structure, context (inputs)
4. Classroom model including rotations/practice/pacing, dosage, use of materials/assessments (classroom model)
5. Behavior – student (indirect)

Although student behavior is referenced in the logic models, it is difficult to measure because it is both a potential mediator and outcome of the targeted interventions. Therefore, student on-task behavior was included as a separate and *indirect* model component.

[61] As described in the logic model, inputs are the resources that support delivery in the classroom and allow it to happen (inputs include professional development, infrastructure, etc.).

Phase 2: Implementation Component Ratings

The overall rating of adequacy of implementation for each of the five components was based on subcomponent and indicator scores. Adequacy was defined as the implementation of intervention components as specified by the developers and as planned by the districts. As described previously, the assumption has been that all model components were specified at the level necessary to promote student improvements in reading skills. Therefore, overall quality of implementation is assessed by the overall rating of adequacy of implementation.

Exhibit 16 presents the identified model components, the subcomponent indicators, the binary codes used for scoring, and the possible score ranges for each component.[62] Each specified subcomponent and indicator was scored based on criteria provided by developers (often, the observed or reported presence or absence). Fidelity ratings for each subcomponent were then assigned using a binary scoring method (a score of 1 or 0). A score range and percentage was then calculated based on these ratings for each component for each teacher. Therefore, percentages and levels of implementation were based on this exhibit of subcomponents and indicators and their associated scores.

[62] Each subcomponent and indicator listed may include more than one item from the data sources used (e.g., observation and survey data) to calculate the rating as previously described.

Exhibit 16. Definition of implementation components and subcomponents

Major Components and Subcomponents	No	Yes	Score Range*
1. Professional Development Participation (attendance)			**0-3**
a. Initial training	0	1	
b. Ongoing workshops, seminars, and/or online courses	0	1	
c. Ongoing mentoring	0	1	
2. Materials/Technology/Assessments			**0-1**
a. Provision/availability	0	1	
3. Classroom Organization/Structure/Context			**0-2**
a. On-schedule for intervention class time	0	1	
b. Teacher-student ratio not exceeded	0	1	
4. Classroom Model Fidelity			**0-8 (READ 180)** **0-7 (Xtreme)**
a.(i) Instructional practices: structured content	0	1	
a.(ii) Instructional practices: research-based instructional methods	0	1	
a.(iii) Instructional practices: responsive teaching	0	1	
b.(i) Dosage of the class: use of rotations	0	1	**(READ 180 only)**
b.(ii) Dosage of the class: pacing for the year	0	1	
b.(iii) Dosage of the class: amount of instructional time	0	1	
c. Use of materials and/or technology	0	1	
d. Use of assessments to inform instruction	0	1	
5. Student Behavior			**0-1**
a. Students on-task (75% or more of the students)	0	1	

* Score range applies to both interventions unless otherwise noted.

For example, four sub-components were identified for the classroom model fidelity component: (1) instructional practices, (2) dosage of the class, (3) use of materials and technology, and (4) use of assessments. One or more indicators were then identified to represent the subcomponents as appropriate. For instance, the instructional practices subcomponent includes three indicators: structured content, research-based instructional practices, and responsive teaching.

In Year 1, classroom model fidelity had four subcomponents (refer to 4a through 4c above). In Year 2, two of these subcomponents, instructional practices and dosage, were further refined and now have three indicators each (refer to the three listed within 4a and 4b above). Finally, READ 180 indicators were added to the provision of materials component to better capture information related to availability of the materials for implementing the intervention.

As described previously, individual ratings were calculated based on the presence or absence of the subcomponent/indicator or based on whether specific criteria (as described in the following pages) were met (1 = yes, adequate; 0 = no, not adequate) and then composite ratings were created (ranging from 1 to 4) for each component. These ratings were in turn used to calculate overall implementation levels, as follows: 1 = no evidence (0 - 24%), 2 = low (25 - 49%), 3 = moderate (50 - 74%), and 4 = adequate (75 - 100%). This level-of-implementation rating system is rudimentary and as such captures the adequacy of implementation only and not the quality of implementation. For example, the amount of mentoring provided may have exceeded the amount specified by the model, yet the rating would still be designated as "adequate." Conversely, if some amount of professional development (e.g., initial training) was received but not the model-specified amount, the initial training subcomponent of professional development would not be given a rating of adequate.

Ratings: Year 1 and Year 2 Changes

Of the targeted intervention teachers trained to teach READ 180 in Year 2, one had been part of the impact or RCT study as a 9th grade READ 180 teacher in the prior year. Therefore, though other teachers had prior experience teaching READ 180, only this teacher had ratings across both study years. To protect this teacher's anonymity, his/her ratings *were not linked* across years in the following READ 180 exhibits. Of the targeted intervention teachers trained to teach Xtreme Reading in Year 2, two had been part of the RCT in the prior year and an additional teacher had prior experience teaching Xtreme Reading not as part of the RCT.

Since there were more than two Xtreme Reading teachers with data across both study years, their ratings *were linked* across years as reported in the following Xtreme Reading exhibits. The following section presents detailed scoring information within each of the five major components.

1. Professional Development Ratings

Three subcomponents were included in the overall implementation rating of professional development: (1) initial training participation for teachers before the school year began; (2) participation in the workshops, seminars, or online courses (e.g., RED courses) offered as planned throughout the nine-month school year; and (3) receipt of ongoing mentoring provided by intervention developers. Information used to assign ratings included district self-report data and district documentation (e.g., computer printouts demonstrating sessions completed for the required RED course, district-provided attendance records, etc.).

The initial training participation ratings were based on attendance per the total number of days as specified prior to the beginning of classes.[63] For READ 180, two initial training sessions were planned (six hours each) as in Year 1. For Xtreme Reading, two initial training sessions were also planned (six hours each). In sum, both interventions planned for 12 initial hours of required professional development.

The participation ratings for workshops, seminars, and the RED online course (for READ 180 only) were based on teacher attendance at these planned professional development offerings throughout the school year. For READ 180, participation in the online course (i.e., completion of seven online sessions)[64] as well as eight seminars (three hours each for a total of twenty-four hours) was planned.

[63] In Year 1 of implementation, three full days of initial training were required for Xtreme Reading teachers, comprising both Xtreme Reading and SIM-CERT content (as described previously). In Year 2, district staff and the developer felt that the needed content for Xtreme Reading could be covered in two days rather than three, so a change was made in initial number of days for professional development prior to the start of the school year. No initial training was required for teachers that had already taught Xtreme Reading in Year 1.

[64] The RED Course facilitator also led debriefing sessions with teachers in order to further their professional development. These sessions were not required as a model component so were not included in ratings.

For Xtreme Reading, attendance in four full-day workshops (lasting approximately six hours per day) was planned (required) following the initial training.

Finally, mentoring ratings were based on the receipt of the total number of monthly mentoring visits.[65] For both interventions, the mentoring visits were to occur once per month for a total of nine visits as per district attendance records.[66] Note that the ratings of participation in professional development do not in any way reflect the nature of engagement of teachers in these sessions, as engagement was not directly measured. However, professional development training sessions are assumed to have included both didactic and experiential elements designed to influence participant engagement and to promote substantive learning.

Adequacy was equated with the *completion of all* training days planned (and thus required) for covering intervention content. In other words, an adequate level of training is reflected by the presence of *all* required components (i.e., a rating of yes) as described. Attendance documentation and online session completion were the sole measures available to assess training participation. Exhibits 17 and 18 present the ratings for READ 180 and Xtreme Reading, respectively.

[65] In Year 1, the mentoring for each intervention began at different points in the nine-month school year based on the initial coordination between districts and developers (as described previously). For Xtreme Reading, mentoring began in October and continued through May for an eight-month period. For READ 180, the mentoring began in December and continued through May for a six-month period.

[66] Note that both of the interventions indicated they conducted additional mentoring visits "as needed" but the ongoing mentoring rating is based solely on the occurrence of the minimum number of visits as required by the models.

Exhibit 17. READ 180: Ratings of professional development participations (attendance) by teacher

READ 180 Teacher	Professional Development % Year 1	Rating Year 1	Professional Development % Year 2	Rating Year 2
1	67%	Moderate	--	--
2	33%	Low	--	--
3	33%	Low	--	--
4	100%	Adequate	--	--
5	67%	Moderate	--	--
6	100%	Adequate	--	--
0*			50%	Moderate
	Mean = 67%			
7	--	--	67%	Moderate
8	--	--	100%	Adequate
9	--	--	100%	Adequate
10	--	--	67%	Moderate
			Mean = 77%	

Note: Implementation levels were defined as: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

Of the three teachers in Year 2 with experience teaching READ 180, all completed the online RED course training. Two of the three participated in each of the nine mentoring sessions offered (per district records) and the remaining teacher participated in eight of nine mentoring sessions. During the mentoring sessions, a classroom observation was conducted followed by a debriefing session.[67] All sessions were reportedly 120 minutes in length (90 minutes plus 30 minutes for debriefing) with the exception of the June visit in which the observation time was cut in half for all teachers.

[67] Three of the five READ 180 teachers taught READ 180 for a second year, and thus were not required (per the plan) to receive initial training and were not required to participate in seminars offered throughout the year. [As noted prior, only one of these three teachers were included in the impact or RCT study sample presented in this report and, therefore, in the exhibits.] These teachers were required, however, to take the online RED course and were expected to receive mentoring in the classroom. These factors were taken into account when tabulating the scores presented (e.g., if a subcomponent such as initial training was not applicable to a teacher, it was not included in that specific calculation).

The two teachers in Year 2 who were new to teaching READ 180 participated in all initial training and completed all online RED course sessions (i.e., seven sessions).[68] As for ongoing mentoring, both of the teachers new in Year 2 participated in almost every mentoring session (i.e., eight of the nine sessions). However, for each of the three training subcomponents—initial, workshops/seminars, and ongoing—teachers needed to meet model requirements of 100% participation (and therefore needed to attend all of the nine mentoring sessions) for a rating of adequate. The missed mentoring session resulted in an overall rating of "moderate" for these teachers.[69] Cohort 2 READ 180 teachers received on average a total of 36 hours of professional development, hours received ranged from 20 to 57 hours. The variation in hours received among teachers was large because the three teachers who had taught READ 180 in Year 1 were not required to attend the initial professional development sessions nor were they expected to attend the four seminars offered to new teachers.

Overall, scores for the receipt of professional development increased from Year 1 (67%) to Year 2 (77%). In the Year 1 survey, only one of the six teachers indicated having received all of the six monthly mentoring sessions (one subcomponent of professional development) as planned for Year 1. Two teachers reported 4 visits; one teacher reported receiving five visits and the remaining teacher did not respond. During this first year of implementation, there were fewer overall sessions because the provision of mentoring assistance did not begin until the end of the calendar year as reported by the districts, resulting in scheduling difficulties. In Year 2, according to district records,[70] the nine monthly mentoring visits began in September 2008 and were provided throughout the academic year as planned.[71] Whereas two of the teachers participated in all nine sessions, three of the five teachers participated in eight sessions

[68] The number of sessions was as per the progress report generated by Scholastic and provided to evaluators by the district.
[69] In Year 2, districts expressed concern regarding the rates of participation in professional development activities as reported by teachers. In Year 1, districts indicated that, per their own documentation, all teachers in the study received all of the planned monthly mentoring visits from the developers. Evaluators acknowledge it is possible teachers under-reported the receipt of monthly professional development visits when responding to the teacher survey. Districts have since requested summaries of mentoring visits (by teacher) from the developers.
[70] In Year 1, survey items were used for scoring; in Year 2 district records were used.
[71] Mentoring visits included classroom observations and a debriefing session.

Exhibit 18. Xtreme Reading: Ratings of professional development participation (attendance) by teacher

Xtreme Reading Teacher	Professional Development % Year 1	Rating Year 1	Professional Development % Year 2	Rating Year 2
1	100%	Adequate	--	--
2	67%	Moderate	--	--
3	33%	Low	100%	Adequate
4	67%	Moderate	100%	Adequate
5	100%	Adequate	--	--
	Mean = 73%			
6	--	--	0%	No evidence
7	--	--	100%	Adequate
8	--	--	100%	Adequate
			Mean = 80 %	

Note: Implementation levels were defined as: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

In Year 2, four of the five Xtreme Reading teachers scored 100% in terms of professional-development attendance. All three of the returning Xtreme Reading teachers received the intended professional development in Year 2; that is, they received both the nine mentoring sessions as well as the one-day Strategy Integration training session required for both new and returning teachers.[72] Of the two new Xtreme Reading teachers in Year 2, one participated in the initial two-day training session, the five days of ongoing training (delivered in four days now in Year 2 as per districts/developers), and the nine mentoring visits. The Xtreme Reading teacher who scored a '0' for adequacy of implementation was hired in October of 2007 so missed the initial training. However, this teacher received eight of the nine mentoring sessions (missing the first session in September 2007), and participated in three of the five days of Xtreme Reading ongoing training.

[72] Although content from the two missed (ongoing) professional development sessions was covered, this teacher was not present during the training session as planned, therefore, the teacher still received a score of 0.

Because this teacher was also teaching a class of Strategic, this teacher participated in Strategic training for two of the five days. That is, since Xtreme Reading and Strategic sessions were offered concurrently, the teacher had to miss two days of Xtreme Reading training. As reported by district staff, this teacher was trained separately during one or more of the mentoring visits in order to make up for the Xtreme Reading content missed in the two sessions.[73] Cohort 2 Xtreme teachers received on average a total of 46 hours of professional development, hours received ranged from 31 to 78 hours. The variation in hours received among teachers was large because the three teachers who taught Xtreme in Year 1 were not required to attend the initial professional development sessions and were only required to attend one of the five ongoing workshops/training sessions.

A less-than-adequate score in the area of mentoring participation for Xtreme Reading teachers was the result of the delayed hiring of a new teacher in Year 2. All other Xtreme Reading teachers were rated as receiving adequate training as defined. The overall higher percentages of adequacy in Year 2 may in part be explained by the difference in the ratings of the mentoring sessions (higher ratings in Year 2 as compared to Year 1). In addition, ratings in Year 1 were based on teacher responses to surveys administered at the end of the school year because attendance records were not yet available, while Year 2 ratings were based on district-provided attendance records. The lag-time between when surveys were administered and when mentoring took place may have resulted in under-reporting for Year 1. Additionally, after a second year of program implementation, districts were better positioned to coordinate, plan, and track professional development activities.

[73] Although content from the two missed (ongoing) professional development sessions was covered, this did not affect the 0 score because of the following criterion: teachers had to be present during a particular day of professional development as defined.

2. Material Provisions Ratings

Teacher survey items were used to rate the overall adequacy of the materials, technology, and assessments made available to teachers. It was determined that self-report data were more reliable for assessing the provision of materials as all required materials were not readily observable during classroom observations. Seven survey items were used to specify whether or not all required materials and/or technology were available for READ 180 implementation in Year 2,[74] whereas one item related to materials was available and used for scoring in Year 1.[75] Evaluators added six items related to the provision of materials to the survey in Year 2 based on more specific information received at the end of Year 1 regarding developer expectations. Exhibits 19 and 20 present the ratings related to the provision of materials for READ 180 and Xtreme Reading respectively.

Exhibit 19. READ 180: Ratings of provision of materials/technology by teacher

READ 180 Teacher	Materials % Year 1	Rating Year 1	Materials % Year 2	Rating Year 2
1	100%	Adequate	--	--
2	100%	Adequate	--	--
3	100%	Adequate	--	--
4	100%	Adequate	--	--
5	100%	Adequate	--	--
6	100%	Adequate	--	--
0*			100%	Adequate
	Mean = 100%			
7	--	--	100%	Adequate
8	--	--	100%	Adequate
9	--	--	0%	No evidence
10	--	--	100%	Adequate
			Mean = 80 %	

Note: Implementation levels were defined as: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

[74] The survey items for READ 180 in Year 2 required teachers to respond "yes" or "no" when asked: *Does your READ 180 classroom have enough…* (1) student books, (2) materials in its READ 180 library, (3) teacher materials, (4) working computers (including headsets and microph1s) to permit each student to rotate through use of the READ 180 software - each day the class meets, (5) working CD players to permit each student to rotate through use of the audio-books - each day the class meets, (6) READ 180 topic CDs in the classroom, and (7) READ 180 materials & technology… *to implement READ 180 effectively?*

[75] The survey item for READ 180 administered in Year 1 was identical to item 7, listed above.

The same four survey items pertaining to the provision of Xtreme Reading materials[76] were included on the Year 1 and Year 2 surveys. For READ 180, four of the five teachers (or 80%) reported having *enough of all of the materials and/or technology* necessary for implementation. In Year 1, 100% of teachers reported having all of the materials/technology necessary. It is important to note that the one teacher scoring '0' for adequacy of materials in Year 2 reported having enough of all materials and technology *with the exception* of READ 180 library materials. While the districts and developers may have distributed all of the materials as required, the response to this particular survey item may be illustrative of a perceived need for more READ 180 library titles in addition to those published and distributed.

All teachers, 100% (five out of five teachers), would have been rated adequate as in Year 1 had the scoring for "provision of materials by READ 180" been calculated using the one item as in the prior year ("Does your READ 180 classroom have enough READ 180 materials and technology to implement READ 180 effectively?"). However, increasing the number of items on the survey to address the provision of a wider array of materials required for implementation was necessary to portray a more accurate representation of this overall component.

[76] The survey items for Xtreme Reading in both Year 1 and Year 2 required teachers to respond "yes" or "no" when asked: *Does your Xtreme Reading classroom have enough of the following materials:* (1) books in the classroom library, (2) student binders, (3) Xtreme Reading posters, and (4) teacher materials?

Exhibit 20. Xtreme Reading: Ratings of provision of materials/technology by teacher

Xtreme Teacher	Materials % Year 1	Rating Year 1	Materials % Year 2	Rating Year 2
1	100%	Adequate	--	--
2	100%	Adequate	--	--
3	100%	Adequate	0%	No evidence
4	100%	Adequate	0%	No evidence
5	100%	Adequate	--	--
	Mean = 100%			
6	--	--	0%	No evidence
7	--	--	100%	Adequate
8	--	--	100%	Adequate
			Mean = 40 %	

Note: Implementation levels were defined as: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

Evaluators will consider using district records, in addition to teacher survey data, to assess the provision of materials should the latter data be provided in Year 3.

For Xtreme Reading, 40% of all teachers reported having *enough of all* materials necessary for implementation. In Year 1, all cases were rated as adequate for this component. In Year 2, three teachers with scores of 0 for adequacy had reported having enough student binders and posters but not enough books for students (one teacher also reported a lack of teacher materials). The reported lack of a sufficient number of books may be explained by the fact that two of the teachers were not new to Xtreme Reading as they also taught Xtreme Reading in Year 1. More specifically, these Year 1 teachers may have been able to move through the material at a faster rate than anticipated in the second year given their familiarity with the intervention. As a result, these teachers may have required an increased number of student books relative to their counterparts who were new to teaching Xtreme Reading.

Additionally, given changes to the pacing calendar for the school year that reduced the amount of time for the behavior and motivation units at the beginning, teachers may have started using the books earlier in the year resulting in the completion of the available book supply earlier than anticipated. The recommended time for the delivery of these units was reduced from a total of four weeks in Year 1 to a total of two weeks in Year 2 (in some instances the units were actually delivered on an as-needed basis per teacher discretion).[77] As in Year 1, teachers were given the option by the developer to add books to the library, provided that reading levels were verified using an online resource approved by the developers of Xtreme Reading.

3. Classroom Organization – Context Rating

Two subcomponents were used to calculate the overall rating of the adequacy of classroom organization and structure: (1) class time allotted in individual school schedules, and (2) observance of teacher-to-student ratios. Classroom observations as well as district-reported information were used to determine both subcomponent ratings. The planned teacher-to-student ratios were one to eighteen for READ 180 and one to fifteen for Xtreme Reading. Exhibits 21 and 22 present ratings for READ 180 and Xtreme Reading, respectively.

[77] Year 1 begins with units addressing behavior (ACHIEVE, Talking Together, SCORE) and motivation (Possible Selves); the behavioral and motivational portion of Xtreme Reading takes approximately four weeks to implement. In Year 2, schools followed an optional calendar, Calendar B, in which this time was reduced by half and teachers were to begin with reading strategies, more specifically, "Word Mapping" during the third day of the third week.

Exhibit 21. READ 180: Ratings of classroom organization and structure by teacher

READ 180 Teacher	Class Structure % Year 1	Rating Year 1	Class Structure % Year 2	Rating Year 2
1	100%	Adequate	--	--
2	100%	Adequate	--	--
3	100%	Adequate	--	--
4	100%	Adequate	--	--
5	100%	Adequate	--	--
6	100%	Adequate	--	--
0*			100%	Adequate
	Mean = 100%			
7	--	--	100%	Adequate
8	--	--	100%	Adequate
9	--	--	100%	Adequate
10	--	--	100%	Adequate
			Mean = 100 %	

Note: Implementation levels were defined as: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

READ 180 requires 90 minutes of intervention class time per day and Xtreme Reading requires 45 minutes of intervention class time per day.

Exhibit 22. Xtreme Reading: Ratings of classroom organization and structure by teacher

Xtreme Reading Teacher	Class Structure % Year 1	Rating Year 1	Class Structure % Year 2	Rating Year 2
1	100%	Adequate	--	--
2	100%	Adequate	--	--
3	100%	Adequate	100%	Adequate
4	100%	Adequate	100%	Adequate
5	100%	Adequate	--	--
	Mean = 100%			
6	--	--	100%	Adequate
7	--	--	100%	Adequate
8	--	--	100%	Adequate
			Mean = 100 %	

Note: Implementation levels were defined as: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

Both interventions were to be implemented as add-on interventions to the districts' regular ELA courses.

4. Classroom Model Fidelity Ratings

Four subcomponents comprised the overall rating of the adequacy of the classroom model: (1) instructional practices[78] including use of structured content, research-based instructional methods, and responsive teaching; (2) dosage,[79] including use of rotations, pacing for the year, and amount of instructional time; (3) use of materials and/or technology; and (4) use of assessments to inform instruction.

All ratings were based on observations; the only exceptions were the use of survey data to rate the use of assessments (as this is not an observable component of daily implementation) and the use of school calendars and pacing guides to rate intervention pacing across the entire year.[80] Two observations were used, when available, to increase reliability. However, the number of times observations were conducted twice in Year 1 was low for both READ 180 and Xtreme Reading (one out of six and three out of five, respectively) as compared to Year 2 (five out of five and four out of five, respectively). Differences between years were the result of complications related to the timing of the grant's receipt (i.e., Year 1 was the start-up year of the grant), teacher absences, and scheduling difficulties. Because of the greater number and use of repeated observations in Year 2 as compared to Year 1, Year 2 ratings are more reliable. However, the scores were based on the observed occurrence of specific subcomponents in *both* instances. That is, when two observations were conducted for a single teacher, a score of 1 was only assigned if the teacher received a score of 1 for both observations.

[78] In Year 1, the 'instructional practices' subcomponent was comprised of only one indicator. In Year 2, this subcomponent was further refined into three separate indicators—structured content, research-based methods and responsive teaching to more accurately capture classroom teaching practices based on information received post-Year 1.

[79] In Year 1, 'dosage' was a subcomponent in and of itself. In Year 2, this subcomponent was further refined into three indicators—use of rotations, pacing for the year and amount of instructional time. Use of rotations was used as an indicator only for READ 180. Although this indicator was specific only to READ 180, it was included because it is the primary method by which the READ 180 classroom model is implemented.

[80] While evaluators were not able to observe teachers more than once or twice, it was assumed that prescribed intervention activities should be observable in every lesson, on any given day.

Ratings for each subcomponent indicator were then added to reach an overall rating for classroom fidelity. The ratings based on observations represent an occurrence of the practice at that point in time. Given that both intervention developers indicated that the first three subcomponents of classroom model implementation (i.e., instructional practices, dosage, and materials) should occur to some degree daily, it would be reasonable to expect that any given daily observation would be a reasonable representation of what regularly occurred in the intervention classes. Exhibit 23 presents the ratings for READ 180.

Exhibit 23. READ 180: Ratings of classroom model fidelity by teacher

READ 180 Teacher	Classroom Fidelity % Year 1	Rating Year 1	Classroom Fidelity % Year 2	Rating Year 2
1	75%	Adequate	--	--
2	100%	Adequate	--	--
3	0%	No evidence	--	--
4	75%	Adequate	--	--
5	0%	No evidence	--	--
6	33%	Low	--	--
0*			25%	Low
	Mean = 47%			
7	--	--	63%	Moderate
8	--	--	88%	Adequate
9	--	--	63%	Moderate
10	--	--	75%	Adequate
			Mean = 63 %	

Note: Implementation levels were defined as: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

Based on the four subcomponents (i.e., instructional practices, dosage of the class, use of materials/technology, and use of assessments), a majority of READ 180 teachers (80% or four out of five teachers) were implementing with adequate or moderate fidelity in Year 2. One teacher implemented the intervention with low fidelity. These ratings were higher than those in Year 1 during which two teachers demonstrated no evidence of classroom model fidelity.

When examining the single subcomponent corresponding to instructional practices used in the first year, teachers showed more variation in their ratings.[81] However, ratings from Year 1 to Year 2 should be compared with caution as new teachers in Year 2 began teaching during the second year of the grant, at which point the start-up phase for the grant was completed and most of the structural elements for implementation were already in place.

In addition, indicators added in Year 2 were based on information about the interventions provided in the second year of the grant. Most teachers received adequate ratings in the instructional practices component (including use of structured content, use of research-based instructional methods, and responsiveness to students). That is, teachers covered the skills or content pertaining to one of the READ 180 workshops and used specific READ 180 instructional strategies during teacher-directed activities in small group and/or whole-class settings. In addition, most teachers provided feedback, monitored comprehension, or provided support for acquisition of skills to one or more students.

For READ 180, dosage comprises the use of rotations, pacing for the year, and the amount of instructional time students receive. Of the five teachers, one teacher demonstrated adequacy for instructional time; that is, one teacher allotted the *full* amount of time to the model-specified instruction (with the exception of transition time).[82] Instructional time is defined as teacher behavior that directly or indirectly supports the model (i.e., activities and conversations related to the intervention and/or goals of the lesson). One teacher met this requirement, which is not surprising as it is not atypical for teachers and students to get sidetracked from teaching and learning for one reason or another.

[81] Additionally, had the same indicators (total of four, one indicator per subcomponent) been used to measure fidelity in Year 2 as in Year 1, the overall rating of classroom-level fidelity for the group of teachers would have been 73% (as compared to 63% in year 1), with four teachers rated as "moderate" and one as "adequate." The pattern of increased fidelity of implementation with regard to professional development for this group of teachers still holds. The Year 2 percentage is likely more accurate because of the increased number of measures and increased number of observations used to calculate ratings (increasing reliability).

[82] The rating for instructional time was developed to capture observed occurrences of on-model behavior for a majority of class time (beyond "settling-in" in the initial few minutes of class) versus those receiving the same rating for exhibiting on-model behavior but not for the majority of planned intervention time (as in Year 1).

However, the rotations were central to the READ 180 model and as such it was surprising that one out of five teachers was observed implementing the three basic READ 180 rotations as specified in Year 2.[83] This finding contrasts with Year 1 in which three out of the six teachers were observed implementing these rotations. In Year 1, fewer instances of small-group instruction were observed. In Year 2, many instances of either small-group instruction or whole-class instruction were observed, but not both. For the third and final dosage indicator, pacing for the year, two teachers were rated as adequate.[84] Three of the five READ 180 teachers were on target (i.e., on Workshop 6) during the first observation in February 2008, and four were on target (i.e., on Workshop 9) during the second observation in May 2008. Two out of the five teachers were on target at both points in time and thus two teachers met the criteria for pacing for the year.[85] Many contextual factors that may have contributed to pacing for the year will be elaborated on in the following sections of the report.

While the materials component describes the provision of materials, the actual use of materials is assessed as outlined in the classroom model. In Year 2, all teachers received a rating of adequate for the use of READ 180 materials and adequate for the use of READ 180 assessments to inform instruction. During the classroom observations, students were observed to be using the rBook, reading novels published by Scholastic, or using READ 180 software.

[83] For rotations to be considered adequate, observers should have seen the whole group instruction and all three rotations (small-group teacher-directed instruction, computer rotation and independent reading). The presence of the "wrap-up" activity was not considered in the scoring.

[84] To score pacing, evaluators mapped the weeks based on district calendars received for Year 2 (taking into account holidays, school breaks, etc.) onto the pacing chart in the teacher planning guide. Schools were visited during the week of February 4, 2007 (day 95) and the week of May 1, 2007 (day 154). According to these calculations, teachers would have been on Workshop 6 (to be completed between days 84 to 97) and on or beyond Workshop 9 (to be completed between days 125-145), respectively.

[85] While in Springfield-Chicopee it was expected that all nine READ 180 workshops be implemented within one academic year, this is not true of other Striving Readers projects implementing READ 180. Across the Striving Readers projects, on a national level, yearly pacing varies.

Additionally, all READ 180 teachers reported administering the required assessments including: the Scholastic Reading Inventory (SRI) for diagnostic information (a minimum of three times during the year), and the rSkills tests given after specific workshops to measure acquisition of READ 180 rBook skills (a minimum of five times per year). The following summary exhibits present survey results for each year.

Exhibit 24. READ 180 teacher survey (Year 1): Summary of teacher responses regarding use of Scholastic Achievement Manager (SAM) (n=5)

	Strongly Disagree or Disagree	Undecided	Strongly Agree or Agree
a) SAM data reports help me implement READ 180		1	4
b) SAM reports help me differentiate instruction			5
c) SAM reports help me assess student progress		1	4
d) SAM reports help me group students			5
e) I share information from the SAM reports with school administrators or other school staff	1		4
f) I share information from the SAM reports with parents		2	3
g) I share information from the SAM reports with students			5

Exhibit 25. READ 180 teacher survey (Year 2): Summary of teacher responses regarding use of SAM (n=5)

	Strongly Disagree or Disagree	Undecided	Strongly Agree or Agree
a) SAM data reports help me implement READ 180		1	4
b) SAM reports help me differentiate instruction	1	2	2
c) SAM reports help me assess student progress			5
d) SAM reports help me group students	1	2	2
e) I share information from the SAM reports with school administrators or other school staff	1	2	2
f) I share information from the SAM reports with parents	2		3
g) I share information from the SAM reports with students			5

All teachers responded that they used reports generated by Scholastic Achievement Manager (SAM). In Year 1, there was more variability in terms of use of materials and use of assessments. In Year 2, as reported by teachers, an adequate rating was achieved for both components (i.e., use of materials and use of assessments). When asked about the usefulness of SAM reports in the teacher survey, a majority of the Year 1 and Year 2 impact study teachers agreed that SAM reports helped them to: (1) implement READ 180, (2) differentiate instruction, (3) assess student progress, and, (4) group students. In Year 1, one of the five teachers responded undecided as to whether SAM helped them implement the program and one responded undecided as to whether SAM helped with assessing student progress. For the Year 2 survey respondents, there was more variation in the responses. More specifically, one of the five teachers disagreed with the statement, "SAM reports help me to differentiate instruction" while one additional teacher disagreed with the statement, "SAM reports help me group my students."

Exhibit 26. Xtreme Reading: Classroom model fidelity ratings by teacher

Xtreme Reading Teacher	Classroom Fidelity % Year 1	Rating Year 1	Classroom Fidelity % Year 2	Rating Year 2
1	100%	Adequate	--	--
2	75%	Adequate	--	--
3	75%	Adequate	57%	Moderate
4	50%	Moderate	57%	Moderate
5	0%	No evidence	--	--
	Mean = 60 %			
6	--	--	43%	Low
7	--	--	43%	Low
8	--	--	43%	Low
			Mean = 49 %	

Note: Implementation levels were defined as: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

Of the five Xtreme Reading teachers, 40% (n=2) implemented the models with moderate fidelity in Year 2. These were the same teachers who implemented Xtreme Reading (as part of the impact study) in the first year as well. The remaining three teachers were implementing at levels of fidelity determined to be low.[86] In terms of instructional practices in Year 2, all Xtreme Reading teachers implemented moderately (67% when looking at this component in isolation), but fared better in responsiveness and structured content than in research-based methods. Whereas 100% of teachers were rated as adequate in use of Xtreme Reading materials (e.g., Bluford books, Xtreme worksheets, Xtreme notebooks, etc.), one teacher received an adequate rating for use of assessments to inform instruction.

The overall rating of classroom model fidelity decreased from 60% in Year 1 to 49% in Year 2, in large part due to the lower ratings teachers received in Year 2 for use of assessments. However, caution must be used in the comparison because a somewhat different set of teachers was observed in Year 2. In addition, had the same items been used for rating the use of assessments in Year 2 as in Year 1, the overall rating of classroom-level fidelity for the group of teachers would have been 70% (as compared to 49%), reversing the downward trend in fidelity.

In Year 1, teachers were asked to report how often they used reading assessments (use of assessments is a subcomponent contributing to the scoring of classroom-level fidelity). In Year 1, two teachers were rated as adequate in use of assessments and one teacher received the rating of no evidence. The final two teachers did not respond to the survey question regarding assessments, therefore this subcomponent was not included in the classroom-fidelity ratings for these teachers. The missing data for two teachers resulted in an inflated overall fidelity score for Year 1. In Year 1, there were also difficulties reported by teachers in obtaining proposed assessments and in understanding how to use the assessments that were provided.

[86] In Year 1, for Xtreme Reading, dosage was measured in terms of weekly lesson plans but not in terms of units completed over the course of the academic year. In Year 1, several Xtreme Reading teachers did not cover all the units as planned for the year; however, this was not captured in the Year 1 scores. Evaluators added pacing in Year 2.

In Year 2, the developer sought to add AIMSweb and other tools for additional support, and modified some of the requirements and tools in Year 2.[87] Based on this information, teachers were asked how often they administered end-of-unit assessments, AIMSweb measures, and the Grade, with adequacy, defined as a minimum of one to two times per year for each assessment. All teachers responded to the survey in Year 2; one teacher received a rating of adequate for the use of assessments. When asked about the quality and the utility of intervention assessments on the survey (using a scale in which 1 = among the worst and 5 = among the best), responses were positive overall with some variation. The five teachers who rated the "quality of unit tests for assessing what students know," had responses that ranged from 2 to 4 with a mean of 3.2. Four of these teachers also rated the "usefulness of student assessment results for planning instruction," with responses ranging from 3 to 5 and a mean of 4.

In Year 2, in terms of dosage (pacing for the year),[88] none of the teachers were on schedule as planned per the Option B pacing calendar provided by developers. Contextual issues affecting the pace of instruction over the course of the year included the scheduling in Springfield of ELA and Xtreme Reading during the same block, which created a conflict between the time for implementation of Xtreme Reading versus the time for instruction in standard ELA. Other reported barriers noted in both districts included challenging student behavior and difficulties with classroom management, lack of administrative support for the program (perhaps related to administrator turnover and communication challenges), and reduction of instructional time due to "testing overload."

[87] The use of AIMSweb was introduced in Year 1 but was not extensively used until Year 2 and was later discontinued in Year 3.

[88] Schools were visited during the week of February 4, 2007 and the week of May 1, 2007. Based on the "Option B" pacing calendar selected by Springfield-Chicopee, we would expect to observe Visual Imagery during week 21 (the week of the first observation) and Inference Strategy during week 34 (the week of the second observation) in order for the pacing to be on target for the year.

5. Student Behavior Rating

One subcomponent was used to rate on-task student behavior using observation data. However, the indicators used to rate student behavior differed by intervention. If most of the students in an observed class (75% or more) were not disruptive and appeared to be exhibiting on-task behavior for the majority of class time, teachers received a score of 1. On-task behavior includes listening to the teacher, engaging in discourse, using intervention writing materials, using technology as prescribed by the model, and using intervention reading materials. This rating reflects student compliance with teacher directives during class time during the classroom model implementation.[89]

Exhibits 27 and 28 present ratings for student behavior for READ 180 and Xtreme Reading, respectively.

Exhibit 27. READ 180: Ratings of behavior (students on-task) by teacher

READ 180 Teacher	Behavior % Year 1	Rating Year 1	Behavior % Year 2	Rating Year 2
1	0%	No evidence	--	--
2	0%	No evidence	--	--
3	0%	No evidence	--	--
4	100%	Adequate	--	--
5	0%	No evidence	--	--
6	100%	Adequate	--	--
0*			0%	No evidence
	Mean = 33%			
7	--	--	0%	No evidence
8	--	--	0%	No evidence
9	--	--	0%	No evidence
10	--	--	0%	No evidence
			Mean = 0 %	

Note: Implementation levels were defined as: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

[89] Although this rating could be considered to be an indicator of teacher skill (i.e., more skilled teachers are presumably better able to keep students on-task) on-task behavior does not necessarily indicate on-model behavior. For example, in READ 180 students could be working on the computer but not using READ 180 tools. That is, students could be using the Internet for purposes not relevant to their daily lesson.

In general, the overall rating of fidelity for student on-task behavior was low in both years, but less consistent in Year 1. In Year 2, there was no evidence of student on-task behavior at the time of observation. On-task behavior or student behavioral expectations were not as explicitly defined by model providers relative to other components. Therefore, student behavior ratings were considered to be an "indirect" intervention component. Although student behavior is not explicitly linked to teacher practice, these behaviors can affect or mediate intervention outcomes. As a result, this indirect component was observed and rated.

Exhibit 28. Xtreme Reading: Ratings of behavior (students on-task), by teacher

Xtreme Reading Teacher	Behavior % Year 1	Rating Year 1	Behavior % Year 2	Rating Year 2
1	100%	Adequate	--	--
2	100%	Adequate	--	--
3	100%	Adequate	0%	No evidence
4	100%	Adequate	100%	Adequate
5	0%	No evidence	--	--
	Mean = 80%			
6	--	--	100%	Adequate
7	--	--	0%	No evidence
8	--	--	100%	Adequate
			Mean = 60%	

Note: Implementation levels were defined as: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

In general, the overall rating of fidelity for Xtreme Reading student on-task behavior was 80% in Year 1 and 60% in Year 2. One of two teachers included in both study years initially had a rating of adequate but no evidence was observed to rate this component in the second year.

Phase 3: Overall Implementation Ratings

The final phase in establishing an overall level of implementation rating for each of the targeted interventions involved compiling the five component ratings by teacher and indicating the numbers of teachers achieving the highest level (i.e., adequacy). To reiterate, a rating of adequate has been defined as implementation of the intervention at the expected level given model specifications. Composite or overall component ratings comprise subcomponents scores.

READ 180: Implementation Ratings

A summary of ratings for each of the five implementation components for READ 180, as described in the prior section, is presented by school and teacher in Exhibit 29. These components were organized into the two broader categories of intervention for comparison across Striving Readers sites: inputs and classroom model. Student behavior ratings were not reported as a part of the overall implementation ratings for reasons described earlier.

Taking the average of the implementation ratings (expressed as percentages) of the three components considered "inputs" (i.e., professional development participation, provision of materials/technology, and classroom organization/structure), four out of the five teachers received ratings of adequate in Year 2, indicating that an adequate level of implementation was achieved by four teachers. One teacher received a rating of moderate based on teacher-provided information regarding the availability of materials/technology. While this teacher reported having all materials to implement the intervention effectively, he/she reported not have enough materials in the READ 180 student library.

Exhibit 29. Summary of READ 180 component ratings Years 1 and 2 (n=10)

	Inputs		Classroom Model	
Teacher	**Year 1**	**Year 2**	**Year 1**	**Year 2**
	1-3. Average inputs	**1-3. Average inputs**	**4. Classroom model**	**4. Classroom model**
1	Adequate	--	Adequate	--
2	Moderate	--	Adequate	--
3	Moderate	--	No evidence	--
4	Adequate	Adequate	Adequate	Low
5	Adequate	--	No evidence	--
6	Adequate	--	Low	--
7	--	Adequate	--	Moderate
8	--	Adequate	--	Adequate
9	--	Moderate	--	Moderate
10	--	Adequate	--	Adequate

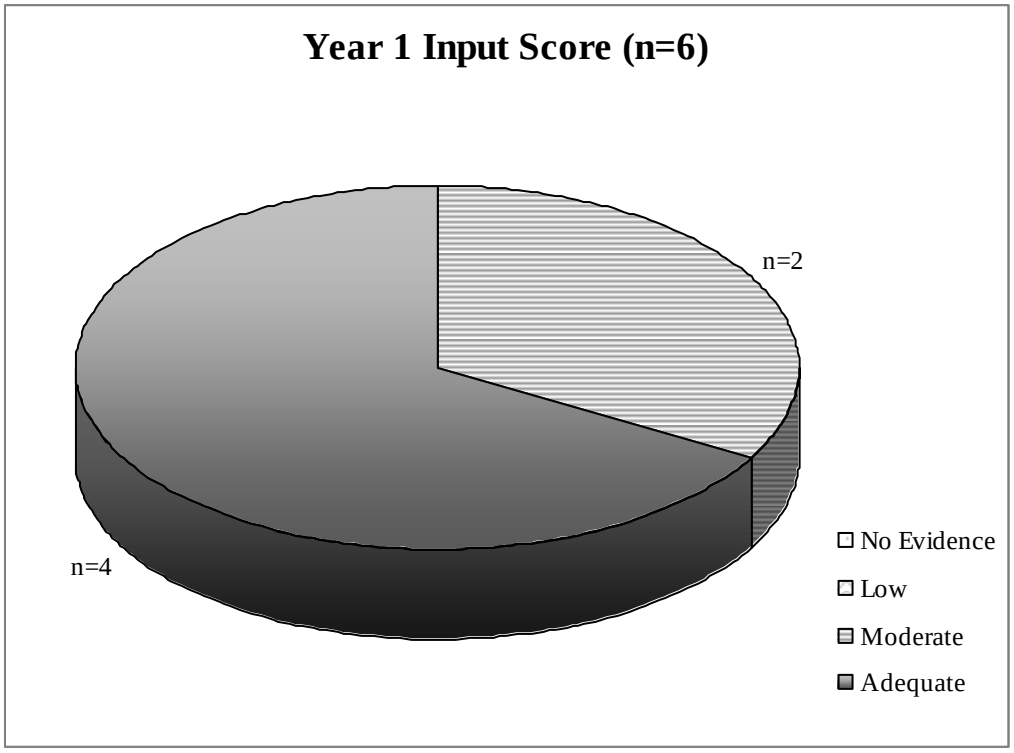
Note: Implementation levels were defined as: 1 = No evidence, 2 = Low, 3 = Moderate, and 4 = Adequate.

For professional development, the third input score, ratings were either moderate or adequate. When Year 2 ratings were compared to Year 1, they were more consistent and positive. As explained earlier in the description of Phase 2, the overall professional development score itself was notably influenced by one of the three items (i.e., mentoring).

For the classroom model, two of the five teachers were rated as adequate, indicating that implementation was achieved. Two were rated as moderate (defined as implementing a majority of model components, a majority of the time), while the remaining teacher was rated as low, indicating that the appropriate level of implementation for the classroom model was not achieved. The teacher rated low was dismissed at the end of Year 2 and replaced by another teacher in Year 3. Most teachers were implementing at a moderate or adequate level in terms of classroom model. In addition, two out of the five (40%) of teachers received ratings of adequate at the end of Year 2 in the implementation of *both* inputs and classroom model (compared to 33% in Year 1).

The remaining three of the five teachers were rated at mixed levels for both inputs and classroom model. Exhibits 30 and 31 present the numbers of READ 180 teachers by each level of implementation for the two categories, inputs and classroom model.

Exhibit 30. Number of Year 1 READ 180 teachers by level of implementation



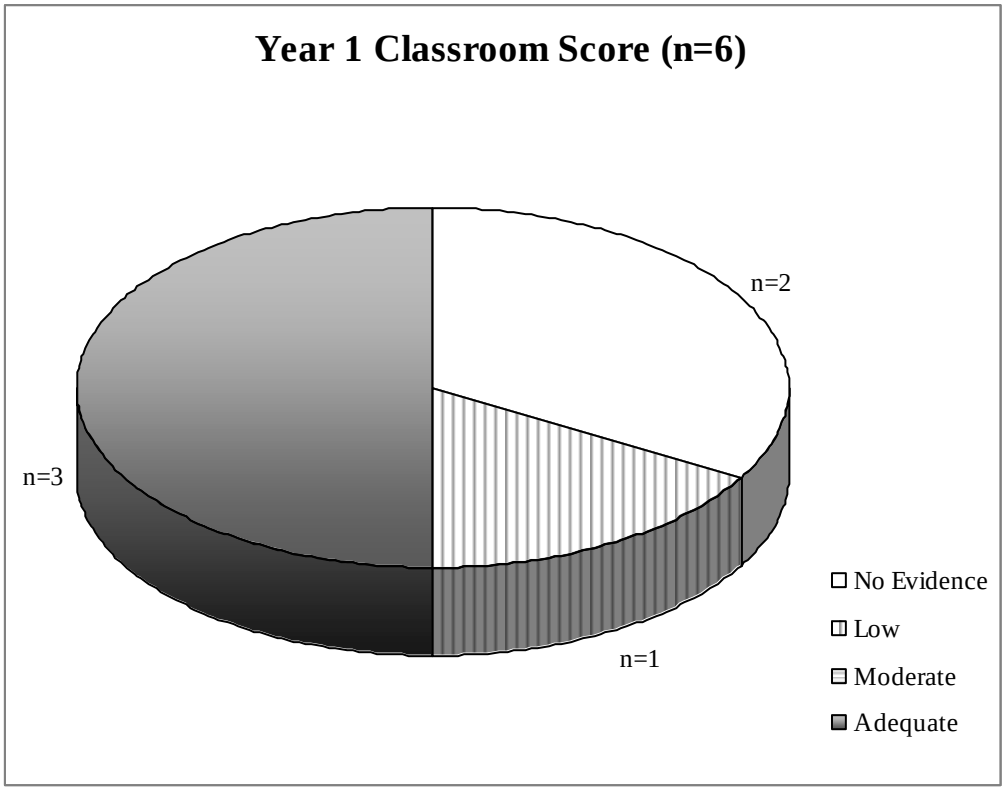
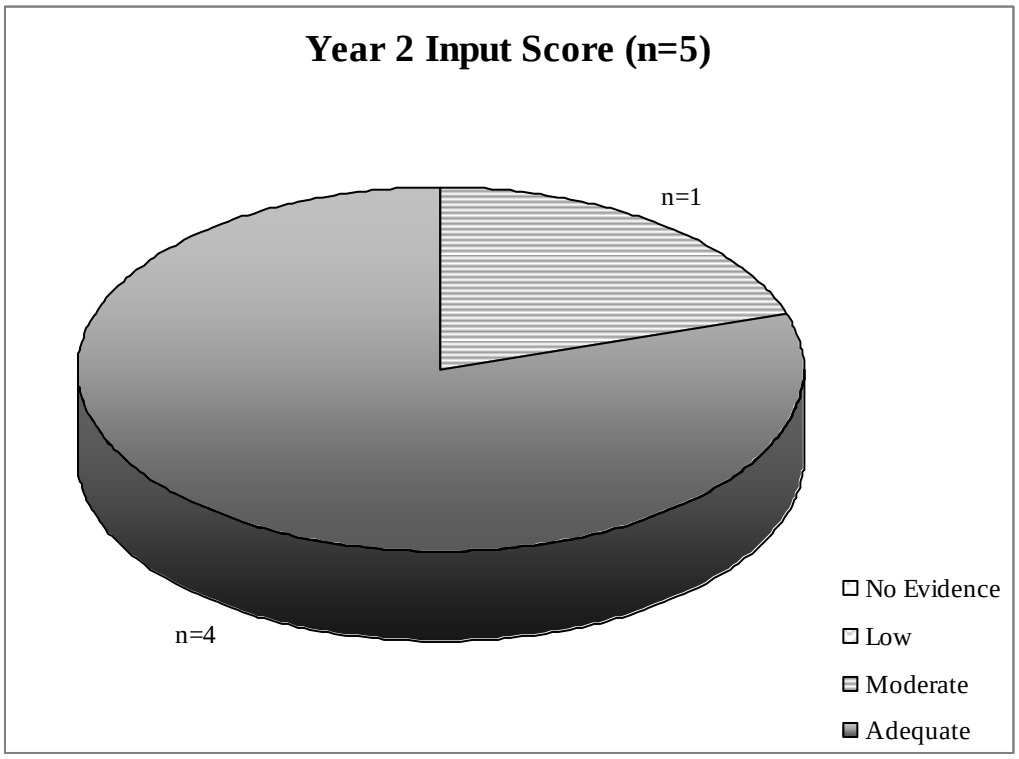
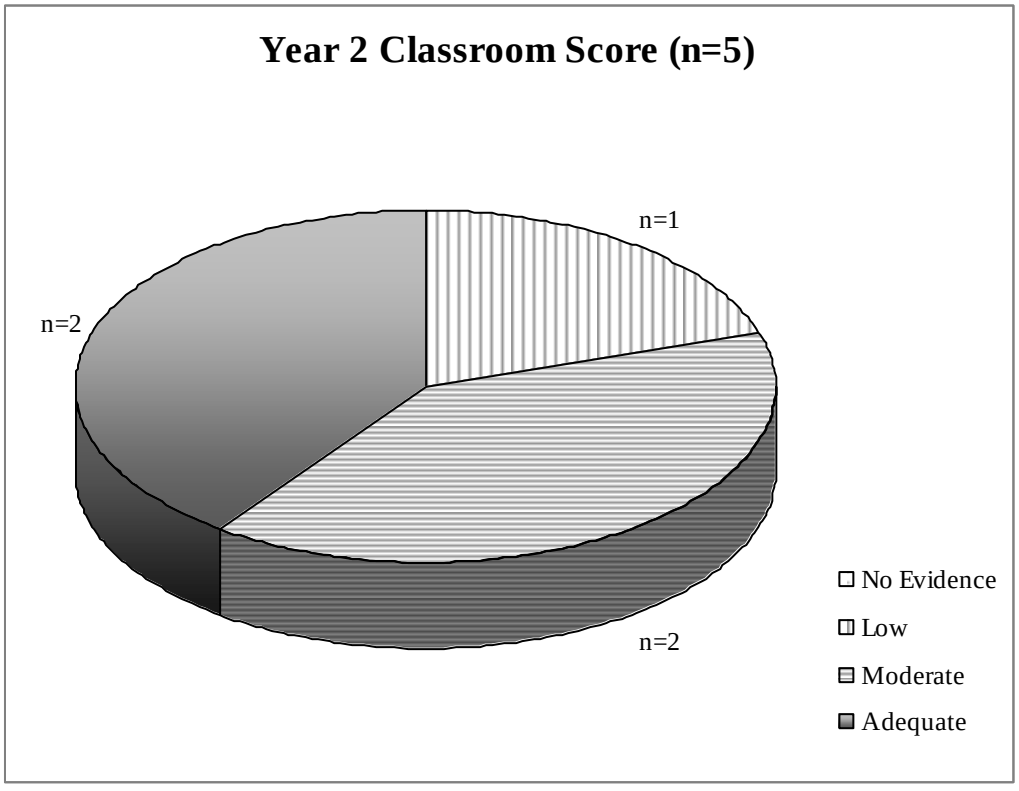


Exhibit 31. Number of Year 2 READ 180 teachers by level of implementation





Xtreme Reading: Implementation Ratings

A summary of each of the five implementation component ratings for Xtreme Reading, as described in the prior section, is presented by school and teacher in Exhibit 32. Components are organized into the two major categories of the intervention: inputs and classroom model. Student behavior ratings were not reported as a part of the overall implementation ratings.

Exhibit 32. Summary of Xtreme Reading component ratings Year 1 and Year 2 (n=9)

	Inputs		Classroom Model	
Teacher	**Year 1**	**Year 2**	**Year 1**	**Year 2**
	1-3. Average inputs	**1-3. Average inputs**	**4. Classroom model**	**4. Classroom model**
1	Adequate	--	Adequate	--
2	Adequate	--	Adequate	--
3	Adequate	Moderate	Adequate	Moderate
4	Moderate	Moderate	Moderate	Moderate
5	Adequate	--	No evidence	--
6	--	Low	--	Low
7	--	Adequate	--	Low
8	--	Adequate	--	Low

Implementation levels were defined as: 1 = No evidence, 2 = Low, 3 = Moderate, and 4 = Adequate.

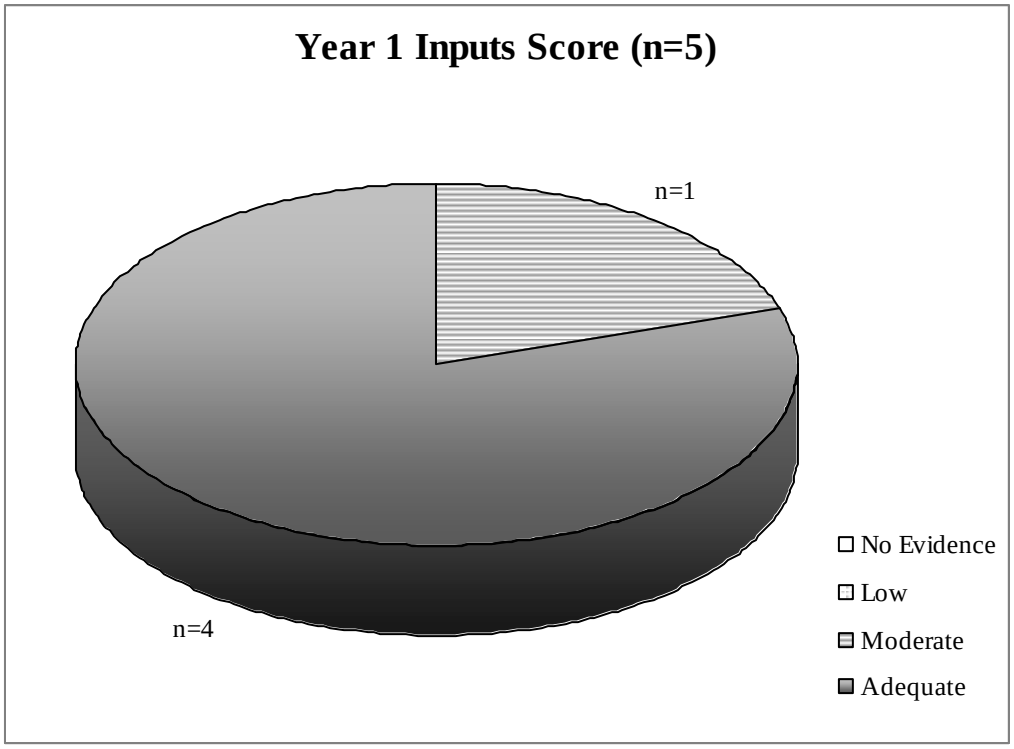
Two of the three model inputs (professional development and class structure) were rated as adequate in Year 2 for four teachers, influencing the overall input ratings. More specifically, two teachers achieved a rating of adequate in Year 2 while four teachers achieved an adequate rating in Year 1. The one teacher who demonstrated no evidence for professional development in Year 2 contributed heavily to the lower overall input score. The absence of professional development for this teacher appeared to be the result of being hired late in the year. Much of the training for this new teacher was designed as "catch-up." In addition, one of the three components contributing to the overall input score (materials/technology) had less consistent ratings for the reasons previously explained.

In Year 1, for the classroom model, three of the five teachers were rated as adequate, indicating the required level of implementation was achieved. The remaining two teachers were rated as moderate and as having no evidence, indicating the required level of implementation for the classroom model was not achieved. In Year 2, two teachers received moderate ratings and three received low ratings. Although there may have been an actual change in classroom level implementation from Year 1 to Year 2, evaluators believe that the use and higher number of more refined indicators addressing classroom level implementation in Year 2 made the scoring more stringent during this second round of data analysis. As stated previously, such measures were added to more accurately capture model fidelity.

Three out of the five teachers received a rating of adequate at the end of Year 1 in the implementation of *both* inputs and classroom model. In Year 2, none of the teachers received ratings of adequate across both categories. Two teachers were rated as moderate, indicating a majority of component indicators were observed. The remaining three teachers received mixed ratings regarding their levels of implementation for both inputs and classroom model.

Exhibits 33 and 34 present the numbers of Xtreme Reading teachers by each implementation level for both the inputs and the classroom model. The exhibits also illustrate variability in Xtreme Reading implementation.

Exhibit 33. Number of Year 1 Xtreme Reading teachers by level of implementation



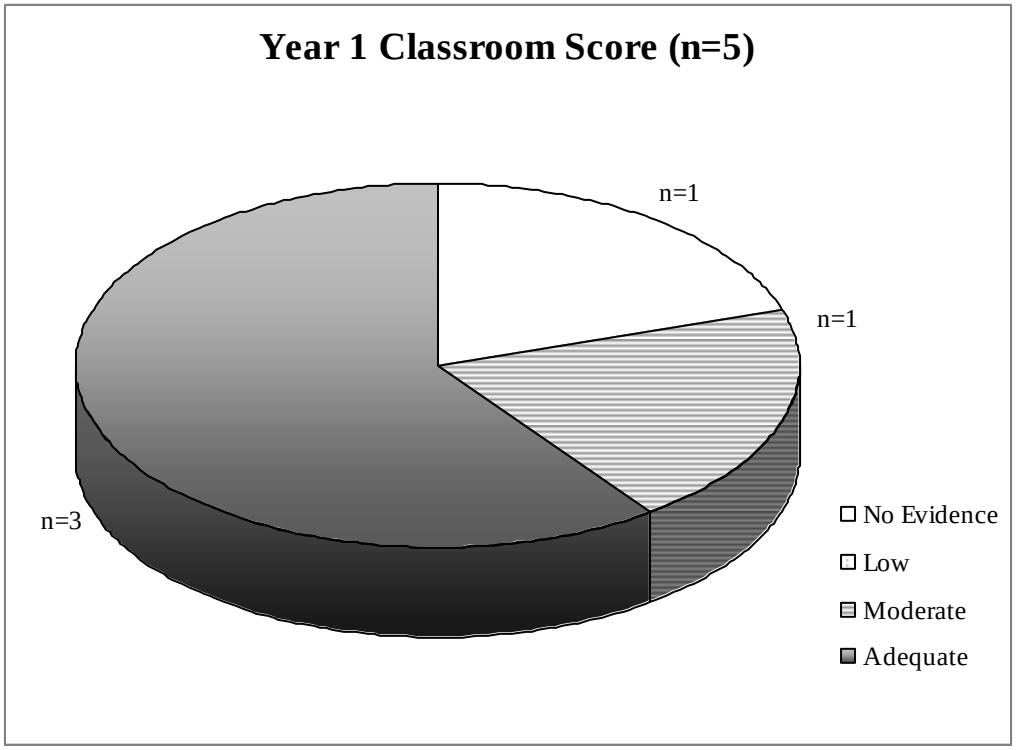
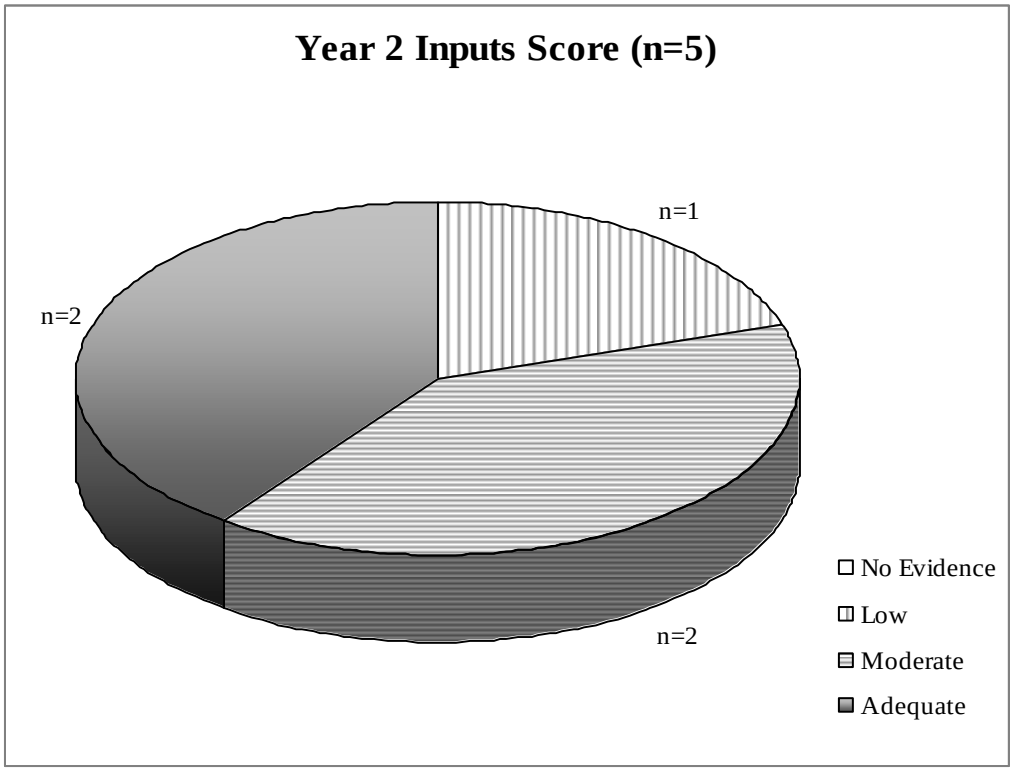
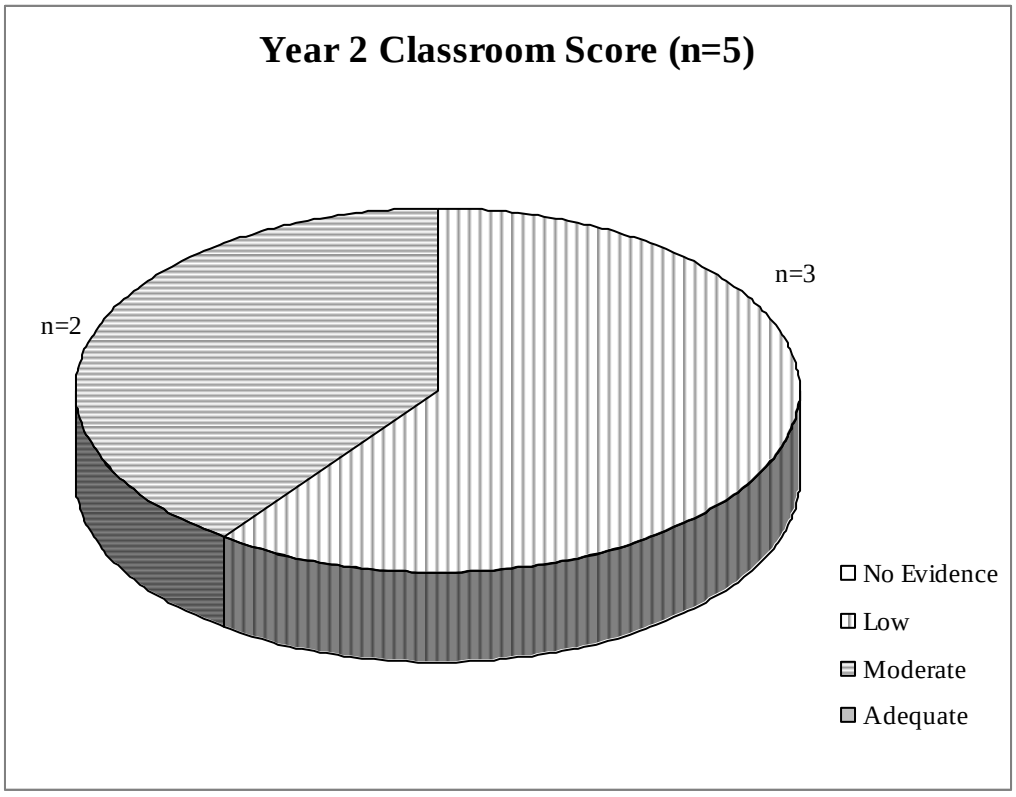


Exhibit 34. Number of Year 2 Xtreme Reading teachers by level of implementation





Intervention Inputs

The overall score for inputs comprises three component scores: (1) professional development; (2) materials, technology, and assessment; and (3) classroom organization, structure, and context. Caution should be used when interpreting these findings to avoid comparing the READ 180 and Xtreme Reading component and subcomponent scores as these have different levels of specificity.

For the professional development component, ratings of adequate professional development for READ 180 in Year 2 (77 %) were higher on average than in Year 1 (67%). The same was true for Xtreme Reading in Year 2 (80%) as compared to Year 1 (73%). The overall higher adequacy percentages in Year 2 may in part be explained by the fact that a greater number of teachers met the minimum standard set regarding the number of mentoring sessions that teachers were to attend.

As per the definition of adequacy used (i.e., 100% participation), 40% of READ 180 teachers (two of five) and 80% of Xtreme Reading teachers (four of five) received an adequate level of training as *planned in Year 2*. In comparison, 33% of READ 180 teachers (two of five) and 40% of Xtreme Reading teachers (two of five) received an adequate level of training *as planned in Year 1*. To reiterate, anything less than 100% participation did not receive a rating of adequate. However, scores of moderate do indicate that a majority of the planned professional development was received.

In Year 2, four Xtreme Reading teachers (two new, two returning) were rated adequate, which is defined as *100% participation in professional development activities*. The remaining Xtreme Reading teacher in Year 2 received a no evidence rating. In Year 1, two out of the five teachers received a rating of adequate. Two of the teachers received ratings of moderate, (indicating that a majority of the training was received), and the remaining teacher received a rating of low. In Year 1, two teachers reported via surveys that mentoring did occur throughout the year but not at the rate of once per month.

The mentoring subcomponent influenced the overall input rating. District documentation and documentation obtained from model developers by the districts was provided to evaluators and reviewed in Year 2.

In both Year 1 and Year 2, based on the aforementioned data sources, 100% of READ 180 teachers and 100% of Xtreme Reading teachers were given the allotted class time for the intervention. Teacher-student ratios were not exceeded in 100% of the intervention classes (READ 180 and Xtreme Reading). Therefore, in all cases, implementation was rated as adequate for this component. Although the allotted time was scheduled for the intervention as specified, the time may or may not have been fully utilized in all cases. At the vocational-technical school, for example, there were numerous challenges faced in the daily implementation plans due to a biweekly academic schedule in both Year 1 and Year 2 of the grant's implementation (refer to the discussion regarding this school in the implications section).

Intervention Classroom Model

The overall score for classroom model comprises four subcomponent scores: (1) instructional rotations and/or practices inclusive of pacing, (2) dosage, (3) use of materials and/or technology, and (4) use of assessments to inform instruction. These items were equally weighted as no developer guidance was provided regarding each subcomponent's importance relative to the overall classroom implementation model. On average, 50% of the READ 180 teachers (three out of six) and 60% of Xtreme Reading teachers (three out of five) were observed to be implementing with adequate classroom model fidelity in Year 1; 63% and 49% in Year 2, respectively. Twenty percent of the Xtreme Reading teachers (one out of five) were observed to be implementing with moderate fidelity in Year 1. Of the five Xtreme Reading teachers in Year 2, 40% (n=2) implemented the models with moderate fidelity. The remaining teachers were observed to be implementing the classroom models with low or no evidence of fidelity.

The classroom model for each intervention differs in practice and prescription. Each intervention is based on its own theoretical model which outlines the ways in which improvements in student reading skills can be attained. The classroom model subcomponents reflect the structure of the interventions and as such differ from one another. For example, the first subcomponent contributing to the overall classroom model rating, instructional rotations and/or practices, was defined with more specificity for READ 180 than for Xtreme Reading because of the manner in which each program has designed its instructional approach. Even with additional specificity, the foundational items/indicators contributing to subcomponent scores were not equivalent for the two programs and so the overall classroom model scores must not be considered equivalent.

As described previously, implementation levels from Year 1 and Year 2 for each intervention may in part be due to the refined specificity and number of measures used to assess classroom level fidelity. Whereas four out of the five teachers were rated on average as implementing at required levels in Year 1, none of the teachers received this rating in Year 2. In Year 3, the same measures will be used as in Year 2, allowing direct comparisons over time of classroom level change within each intervention. Caution should be exercised when interpreting input model scores as each of these is influenced by a varying number of subcomponent scores and because the definition of adequate, as applied to each indicator score, often differs.

IV.C. Implications for the Targeted Interventions

The goal of the implementation study was to present a broad picture of the overall level of implementation for each of the targeted interventions, READ 180 and Xtreme Reading. The districts planned and implemented the two targeted interventions in all grades (nine through twelve), but interventions were implemented in a randomized controlled trial (RCT) only in the ninth grade. Implementation was assessed for each RCT study year and findings provide contextual information to inform the interpretation of the results from the impact analyses. The implementation study entailed assigning ratings for adequacy based on the presence of observed model components as defined by the interventions' developers and the districts prior to implementation (i.e., the model as planned).

While implementation results in this report are presented for both the targeted interventions, a comparison of the interventions is not intended. A summary of the findings is presented in the following pages.

Overall Implications (What Ratings May Not Illuminate)

READ 180

Satisfaction with Professional Development

In addition to ratings of participation in professional development activities, evaluators collected teacher perception data (via an online survey) on satisfaction with the training and support provided by READ 180 developers. More specifically, teachers were asked about the usefulness, amount, and quality of professional development.

According to the teacher survey administered to ninth grade teachers in Year 2 (five of five), most teachers (four of five) agreed that on-site mentoring by READ 180 professional developers helped them to implement READ 180 in their classroom. One teacher disagreed with this statement, saying coaching sessions were not helpful. Year 1 teachers responded similarly to the same item when surveyed in Year 1; three agreed that the on-site mentoring was helpful for classroom-level implementation of the intervention and one disagreed with this statement.[90] This general pattern was observed in teacher interview data as well.

In terms of the *amount* of professional development delivered (including initial training, seminars, online RED course, and in-class mentoring), most of the teachers were satisfied with the professional development provided. Four of the five teachers responding to the Year 1 survey indicated that the amount delivered was sufficient. However, one teacher felt that there was too much professional development. In Year 2, all five respondents agreed that the amount of professional development received in Year 2 was sufficient.

When asked about *quality* of professional development, four of the five READ 180 teachers responding to the Year 1 survey agreed that the professional development offered was of *high quality*.[91] However, the number of teachers reporting satisfaction with the quality of READ 180 professional development decreased in Year 2. More specifically, three of the five teachers in Year 2 agreed that professional development offered was of high quality and two disagreed with this statement.

Barriers to Implementation and Pacing

In Year 1, teachers mentioned scheduling challenges (especially in the vocational-technical school) and mandatory testing as barriers to classroom-level implementation because these reduced instructional time.

[90] In Year 1, three of the five teachers agreed that onsite coaching helped them implement the model, one was undecided, and one disagreed.
[91] One of the six READ 180 teachers in Year 1 did not respond to the survey.

In Year 2, these challenges were cited again via interviews. In addition, teachers noted concerns about the accuracy of student placement (i.e., not all students "belonged" in the class) and about poor student attendance. Teachers also indicated that the occurrence of student advisory periods once per week cut into class time (this period was part of the Small Learning Communities (SLC) grant and adjustments were later made to meet SR instructional time requirements).[92] A few teachers mentioned that they struggled with classroom management, and as a result, had difficulty with transitioning between READ 180 rotations. Finally, based on district records, the average number of teacher absences in Year 2 was 4.5 days (with days of absence ranging from 2 to 8.5 days). All of these factors may have contributed to pacing challenges during the year.

Adaptations

Classroom management difficulties (leading to challenges in transitioning between rotations) and small class sizes (making it difficult to distinguish between whole-class and small group instruction) were offered as explanations for the lack of adherence to the READ 180 instructional segments and rotations in Year 2.

Three of six teachers in Year 1 were observed implementing all instructional segments. In Year 2, instructional segments were implemented to a lesser degree. That is, only one of the five teachers was observed implementing all of the instructional segments (whole-class, small-group, independent reading, and the computer rotations) during the two classroom observations conducted.[93] Given that the whole-class segments and rotations are central to the program, it is important to note this adaptation.

[92] Districts reported an advisory period began in January of 2008 in only one school. Since the READ 180 class was year-long and scheduled the advisory period conflicted, but this oversight in the master schedule was rectified.
[93] In Year 1, only one of the six teachers was observed twice, the remaining five teachers were observed once.

Furthermore, in Year 1, all of the teachers surveyed and/or interviewed (five of six) reported that they made adaptations to the model, most often through the addition of materials. Specifically, adaptations included additional texts, MCAS or assessment preparation materials, vocabulary, writing (including John Collins, one of the district-wide programs), supplementary reading, and other ELA class materials used in the school. Per interview and survey data, the same adaptations were made in Year 2.

Specifically, in Year 2, all five survey respondents reported making "small" adaptations to the activities suggested in the READ 180 Teacher's Manual at a minimum of one to two times per week.. The judgment regarding the magnitude of the adaptation(s) was left to the teacher. When asked about substantial changes in the same way, two teachers reported making what they considered to be "substantial" changes, one to two times per week, while the remaining three teachers did not report making substantial changes to the READ 180 Teacher's Manual during the course of a "typical" week. In Year 1, the number of teachers reporting small and substantial adaptations mirrored those reported in Year 2.

The significance of teacher-initiated adaptations to the program is difficult for evaluators to assess. Teachers indicated they made these adaptations to accommodate for such things as students' ability and needs, inaccurate placement, and competing district mandates/programs focused on literacy and high-stakes testing. The initial and primary reason teachers report adaptations may reflect their approach to appropriate practice in teaching in general, rather than as a barrier to implementation or issue with interventions. This interpretation was supported by the fact that teachers previously cited their remaining explanations for adaptations as barriers. As stated earlier, developers provided additional titles at higher reading levels to encourage differentiation and more variety. However, this modification may constitute a change from Year 1 to Year 2 per district and developer specifications, rather than an adaptation originating from individual teachers.

Xtreme Reading

Satisfaction with Professional Development

Regarding professional development, four of the five teachers received all eight of the mentoring sessions in the planned time period. While all (except for one teacher hired in October) attended every workshop, three teachers reported that the *amount* of professional development delivered for Year 2 was insufficient. More specifically, these Xtreme Reading teachers stated that the professional development offered by the SIM team was inadequate and that they wished they had more training opportunities throughout the year. One teacher indicated feeling unprepared to teach Xtreme Reading for reasons such as not knowing how to assess reading comprehension or not understanding how to use various resources. While all three teachers responding to the survey in Year 1 agreed that the professional development was of high *quality*, two stated that the amount was sufficient and the remaining teacher was undecided. Therefore, more teachers in the second year expressed the need for additional training (two teachers were required to teach Xtreme and Strategic, the intervention delivered in a second year for eligible students). In Year 1, three of the five teachers responded to the survey, which may account for some of the differences in responses between years. However, Year 1 and Year 2 results were consistent with those from the interviews conducted with all teachers.

Barriers to Implementation

In addition to scoring receipt of Xtreme Reading materials, evaluators collected data on the timing of the distribution and the quality of Xtreme Reading materials. In Year 1, teachers were asked via survey about implementation inputs—materials, professional development, classroom organization. All five Xtreme Reading teachers reported the receipt of all materials. However, based on district reports of the timing of the receipt of materials, teachers may not have known the full complement of what they were to receive.

Districts reported that some materials were provided later in the school year than anticipated (rather than at the beginning of the year) and that some materials were further developed after having been distributed to teachers. In Year 1, all three respondents to the teacher survey commented that the student binders were too voluminous and that the materials needed to be condensed. One teacher commented that this was true of the teacher materials as well. In Year 2, one teacher commented in an interview that the materials were disorganized and were not delivered in a timely matter. Teachers interviewed in Year 2 noted that the student binders and teacher manuals had been reorganized.

When SIM developers were asked by evaluators and the SR district implementation team to characterize the changes initiated at the end of Year 2, SIM developers stated that the changes did not pertain to substance but rather to the structure of the lesson plans (e.g., guided reading was placed at the end of the lessons as opposed to the middle, length of time dedicated to the initial start-up activity was reduced, instructions to the teacher were simplified, and no specific references to book pages were included).

Additionally, Year 2 teacher survey results reflect a wide range of responses to items that asked teachers to rate 2007-08 materials in comparison to other curricula used. While two teachers indicated that the organization of the teacher manual and ease of following the lesson plans was "among the best" curricula they had ever used, three teachers responded that these same materials were "among the worst" they had used. In addition, teacher background, number of years of teaching experience, and exposure to other curriculum may have influenced this perception.

Barriers to Pacing

In Year 1, teachers reported that pacing expectations were unrealistic and that there was not enough time in the school year to complete the implementation of all strategies. Other factors identified as influencing pace included the rate at which testing occurred, student absenteeism, and the block-scheduling of the vocational-technical high school which operated on alternate academic weeks.

Additionally, teachers reported concerns regarding student placement and a lack of planning time as barriers to implementation. As described previously, concerns regarding communication and difficulties in implementing the screening and verification processes in one district may have contributed to some of these concerns.

In Year 2, some of the same barriers were reported as in Year 1, with the addition of classroom management difficulties and a lack of awareness by administrators and guidance counselors about the program (which may in part be explained by the rate of administrator turnover during Years 1 and 2 of the grant). A total of fifteen administrators have occupied the ten school administrator positions in two years, including new principals (three of five schools). A number of teachers stated that unrealistic expectations for covering the ELA curriculum (particularly in one district) may have also added to pacing difficulties. In Year 2, pacing across the year was assessed (this subcomponent was not assessed in Year 1 because this was not available). None of the teachers received adequate scores based on this assessment, perhaps demonstrating the impact that structural factors (e.g., scheduling and administrator support) may have on implementation. Further, during the interviews conducted in May 2008, two teachers stated that they would cover the entire curriculum by the end of the year, whereas the other two teachers indicated they did not think they would be able to cover the "Integration" strategy, the last Xtreme Reading strategy to be covered in the year. As stated earlier in this report, the SR district implementation team and developers addressed the issue of administrator knowledge of the interventions by holding informational meetings with new administrative staff. Finally, teachers missed an average of 11.8 days in Year 2 (ranging from 8 to 16 days). Because the sum total of teacher absences was between two and three weeks, this may also have influenced pacing of instruction.

Model Adaptations

In Year 1, evaluators reported that in-class adaptations to the model made by individual teachers included the addition of more writing and vocabulary activities. All three of the survey respondents in Year 1 reported making small as well as substantial changes to their lesson plans at a minimum of one to two times during a typical week.

In Year 2, all five teachers reported making what they considered to be small adaptations at a minimum of one to two times during a typical week, while three teachers made what they considered substantial changes at a minimum of one to two times per week. In Year 2, teachers provided the following examples, during interviews, of adaptations they had made:

- using ELA-related material (e.g., literature);
- requiring students to complete additional writing assignments;
- using Collins Writing strategies;
- reducing the time spent on Xtreme Reading stages such as guided and paired practice in order to transition more rapidly into independent practice; and,
- creating integrated worksheets that combined questions and learning tasks on various Xtreme Reading strategies.

In Year 1, a similar pattern of adaptations was reported. It is not clear based on the information available whether or not the reported adaptations meaningfully change or affect on-model delivery. It is important to note that in Year 2, several adaptations were planned and made by the developer. These changes, which occurred at a macro-level and were thus not teacher-initiated, included: (1) modifications to the pacing calendar (minimizing the time spent on specific socio-behavioral units such as Achieve Expectations, Score, and Talking Together); (2) alterations to the materials; and (3) the introduction and subsequent discontinuation of AIMSweb as a formative assessment tool.

In Year 1, final plans for Xtreme Reading were settled immediately prior to implementation and included the developer requirement that the whole-school intervention training (SIM-CERT) be delivered to Xtreme Reading teachers as well as to the ELA teachers of Xtreme Reading students. However, each district employed targeted teachers differently: Springfield Public Schools' Xtreme Reading teachers taught interventions *and* ELA to the same group of students back to back; Chicopee Public Schools' Xtreme Reading teachers taught only intervention strategies to their students (who then received ELA instruction with another teacher).

In Year 2, SIM-CERT training was not delivered to Xtreme Reading teachers, a decision made by developers in order to allow teachers to dedicate more time to Xtreme Reading strategies. Both the manner in which the two districts used Xtreme Reading teachers and the discontinuation of SIM-CERT training for Xtreme Reading teachers present difficulties in the interpretation of impact analyses as these circumstances could confound study results. The 90-minute block scheduling in Springfield meant that, for Xtreme Reading, there was an additional 45 minutes of instruction in the class period. As a result, students in the Xtreme Reading classes remained for the second 45 minutes to receive standard ELA.

Additional Context (What Ratings Will Not Illuminate)

As previously stated, the interventions were not equivalent and results, including ratings, should not be compared. However, the study of the implementation of two unique interventions affords an opportunity to identify common contextual patterns and potential barriers within the schools and districts that may be unrelated to the interventions themselves but that could influence results. For example, districts as well as individual schools face different challenges. Observed variation may result from such challenges—above and beyond the challenges of implementation of the intervention models. If observed variations appear systematic within districts or schools, it has implications for the analysis and interpretation of impacts. If implementation barriers were systemic within a school or district, these issues would presumably arise in the implementation of either intervention.

Proposed post-award revisions to the Striving Readers district implementation plan, though intended to further ensure consistent implementation across the Springfield and Chicopee schools, did not resolve all challenges. Moreover, implementation in the first year of the grant was challenging in part because it was the beginning of a new initiative; however, key barriers to implementation noted in the first year continued into the second year. Several of these barriers are described in detail below.

ELA and Xtreme Reading Instruction: Springfield

District implementation plans were specified in order to meet developer requirements and to maintain consistency across districts. These implementation plans stipulated Xtreme Reading students would receive 45 minutes of supplemental reading instruction in addition to the 45 minutes of regular English Language Arts curriculum. The ELA instruction was originally to be provided by another teacher in a mainstreamed classroom, given the targeted interventions were considered add-on instruction.

However, Springfield used the same teachers to provide both the Xtreme Reading intervention and ELA instruction in a single class period due to the realities of general block scheduling (i.e., 90 minute blocks). Because the same teachers in Springfield were teaching both Xtreme Reading and ELA, they were instructed to devote 45 minutes to Xtreme Reading and 45 minutes to ELA separately, as was the original plan for alignment in the curriculum across districts.

The decision to implement with the same teachers during a single class period in this manner had several resulting consequences. First, using the same teacher to teach both classes within the block meant that the students participated in an ELA class with a lower teacher-to-student ratio than other ELA classes including the control class. That is, when these Xtreme Reading students remained in the classroom to receive ELA instruction delivered by their Xtreme Reading teacher, they remained in a reduced size class (15 maximum) as compared to their non-Xtreme Reading counterparts who received ELA instruction in a typical size class (i.e., with more than 25 students). Second, because ELA was provided by the same teacher back-to-back with Xtreme Reading, it was often provided in a blended manner. That is, dosage was confounded by the fact that teachers often blended the ELA and Xtreme Reading materials and content. They appeared to do so for several reasons: (1) it was difficult to stop and start at a precise time within the same block, (2) it made sense pedagogically, and (3) it was encouraged by SIM trainers to meet the needs of teachers.

Thus, the ELA curriculum that the Xtreme Reading targeted students received in Springfield was not "business as usual" and the Xtreme Reading intervention was not purely the add-on instruction that was planned or that was implemented in Chicopee, particularly when blending occurred.[94]

District interviews confirmed that communication with all stakeholders in general had been challenging given the turnover of staff, though the SR district implementation team had communicated to teachers their duties in implementing on model. In Year 2, the team had developed and disseminated a handbook for teachers entitled, "Striving Readers Teacher Handbook." This handbook contained a description of the interventions, evaluation information, policies for Striving Reader teachers (e.g., restrictions related to testing, communication, protocols), important dates, and a Striving Readers staff directory. Although the SR district implementation team communicated directly to teachers and administrators that ELA and Xtreme Reading should not be blended, there appeared to be no firmly established school or district policy to this effect, nor was there an accountability mechanism by which to ensure that implementation occurred as planned, at least until later in Year 2. Evaluator observations and interviews with teachers later in Year 2 suggest Springfield teachers had, in fact, received clarification on these issues and were attempting to implement on-model and not blend Xtreme Reading strategies with ELA instruction. While steps were taken to establish additional policies to facilitate communication about Striving Readers, measures to ensure classroom-level implementation were also being developed.

ELA Requirements: Springfield

In Springfield's vocational-technical school, teachers encountered ambiguity regarding how to both address ELA standards and implement the intervention model in their teaching given the school's unique technical configuration.

[94] The potential for blending ELA existed for READ 180 as well, but did not appear to occur primarily because READ 180 requires 90 minutes of instruction time so had a distinct block-period separate from ELA.

Specifically, this school utilizes a block schedule, on a bi-weekly rotation, in which students participate in vocational training during one week, and academic instruction during the next. The block scheduling as described meant that only half of the academic time normally available could be used to meet ELA requirements, including MCAS preparation which is provided *in addition to ELA* for students in need throughout the district. Because of the intricacies of block scheduling and reduced academic time, meeting dosage requirements as proposed was difficult in practice, particularly in the case of READ 180 which required 90 minutes versus the 45 minutes required for Xtreme Reading. Students were scheduled to receive intervention time as planned, however, as a result of the complexities mentioned above, they were scheduled to receive only MCAS preparation (a course to be offered in tandem with and not in lieu of ELA).

Because intervention dosage requirements for scheduling were met in ninth grade, the SR district implementation team did not initially recognize the problem. When the team inquired as to whether MCAS preparation met the ELA standards, the team was provided with positive assurances that students received ELA requirements (i.e., that MCAS preparation met ELA requirements). However, this was later confirmed not to be the case.[95] One source of confusion may have been the fact that Xtreme Reading teachers would have had 45 additional minutes in the 90minute block in which they delivered the Xtreme Reading intervention to include additional ELA instruction, whereas the entire 90 minutes was required for the READ 180 intervention. Although the SR district implementation team had communicated the requirements of intervention delivery, teachers received conflicting messages regarding the ways in which the required state and district standards of ELA were being met via the MCAS preparation course or how they were to meet them in the case of READ 180. As a result, implementation remained complicated.

[95] Key administrators left positions during this time which could have contributed to the confusion and dissemination of inaccurate information.

In Year 1 and at the outset of Year 2, it was reported that teachers chose to meet standards they felt were not being met (though this was not directly observed). However, later in Year 2, teachers reported on-model implementation.[96]

The SR district implementation team also noted formal classroom observations were not yet being conducted given the challenges faced in Year 1 implementation but that plans and processes for observations were being put into place. Thus, during their informal and less frequent class visits, the SR district implementation team may not have observed that teachers were implementing differently or including ELA requirements.. Later, however, the SR district implementation team reported that teachers had independently modified the READ 180 intervention model in order to compensate for the lack of ELA curriculum. This finding was corroborated by evaluator data from teacher observations, interviews, and surveys in both Year 1 and Year 2. Teachers reported that they compensated by adding ELA material so that they could better meet their students' needs.

Assurances were made that dosage requirements would be met as planned in the second year, and the same planned dosage time was scheduled. However, the inherent scheduling difficulties remained in Year 2 with respect to the delivery of the interventions and with respect to ELA instruction; targeted intervention students were again enrolled in MCAS preparation as they were in Year 1. As a result, students in this school experienced a lack of on-model delivery for READ 180 and an intended business-as-usual ELA curriculum that was substantively different from their peers in both content and context in both school years. In addition, the "blending" of ELA and Xtreme Reading also resulted in a lack of on-model delivery for the Xtreme Reading intervention throughout Springfield's schools (as described in the previous sections).[97]

[96] This was not the case for implementation in the upper-grades, where teachers were observed to be off-model in the first and second years.

[97] In the first year, Xtreme Reading specifications were not as clear as in the second year and as the intervention is generally malleable "blending" ELA curriculum may not have the same consequences to model delivery as it potentially does for READ 180.

Pilot Conversion: Springfield

The vocational-technical school in Springfield faced unique implementation challenges, as compared to the other Striving Readers schools, because of its bi-weekly academic schedules but also because of its conversion to Commonwealth Pilot School status during Year 2 (and the planning for this conversion in Year 1) of the Striving Readers grant. Many barriers faced in the work of Striving Readers implementation were echoed in the implementation of the Pilot School initiative. Challenges faced in the Pilot School effort were detailed in an evaluation report from the University of Massachusetts (UMass) Donahue Institute (2008), a first-year evaluation of the status of four Massachusetts schools designated "underperforming" by the state Board of Elementary and Secondary Education. This report provided additional context for Striving Readers regarding the predicaments this newly reconstituted Commonwealth Pilot School faced.

As noted in the Donahue report, reconstituted Commonwealth Pilot Schools are often small (n~400), whereas this vocational-technical school is large (n=1,472). The district therefore proposed building on existing Smaller Learning Communities (SLCs) in this school and creating five distinct "schools-within-a-school" based on this framework. The vocational-technical school did not experience the same degree of turnover as other "underperforming" schools in the district, keeping both its principal and most of the faculty, because vocational staff could not readily be transferred to other district schools. The Donahue report also noted staff surveys indicated a slight majority at this vocational-technical school (57%) were in favor of the school reform efforts that were underway, though staff appeared to have mixed feelings about the status of these efforts at the time of the Striving Readers survey (note: Striving Readers teachers reported many ongoing and overlapping reform efforts in the Striving Readers survey). After a year of Pilot School conversion implementation, staff described a school that was better able to "make decisions," but that demonstrated little to no growth "…in the areas of curriculum, the use of assessment data, student behavior, and the relevance of professional development" (UMass Donahue Institute, 2008, p.46).

The Striving Readers district team noted (in conversations and in a memorandum provided to the Striving Readers evaluators) conversion to Pilot School status required "massive" restructuring efforts involving planning and implementation which impacted the Striving Readers grant implementation in both Years 1 and 2. In the case of the vocational-technical school, the district had to undergo the conversion in the allotted time or be forced to place school oversight into the hands of the state. Springfield discussed the possibility of requesting a waiver to the state requirement for conversion due to Striving Readers study participation. In the end, factors such as state autonomy and the quick start-up of the conversion process rendered it unfeasible to receive a waiver. Efforts then turned toward eliminating or at least minimizing conversion-related factors that could potentially affect teachers' implementation of the ninth grade targeted interventions. The five separate "schools-within-a-school" presented a problem for the study because it split the school's targeted student population, potentially leading to intervention class sizes that were too small to be considered on-model. To counter this potential obstacle, plans were made to establish intervention classes across the five schools.[98] This plan reportedly could not be fully executed in either year, given across-school scheduling complexities.

Teacher Recruitment and Quality

The recruitment plans put into place differed by district given the fact that each district had its own policies, systems, and practices related to the hiring of teachers. In addition, Springfield did not hire all "new" teachers because it had been in the midst of teacher layoffs. The district made the decision to include some of these laid-off teachers in the Striving Readers teacher pool in order to curtail teacher job losses. Additionally, because recruiting qualified candidates was historically challenging, the district felt the need to broaden its pool.

[98] Evaluators acknowledge affects, if any, of reduced class size could be positive or negative but plans were made to implement as consistently as had been done prior to the conversion of this school to Pilot School status. ELA class sizes did not face the same reductions in class size give they would be large enough even if constituted within each of the newly constituted schools.

Overall, fewer teachers were hired for Striving Readers in Springfield and Chicopee because initial estimates were based on the projected numbers of *all* students reading two grade levels or more below their current grade. Later however, adjustments were made to ensure an equivalent comparison group for both interventions, necessitated by the SIM-Xtreme Reading intervention model specifications prior to implementation, which narrowed the range of striving readers to be identified for study purposes.[99]

The direct assessment of *teacher quality* in a valid and reliable way was beyond the scope of the specified grant evaluation activities. All teachers were randomly assigned to teach either one of the targeted interventions or standard ELA.[100] Random assignment was employed to help ensure that teacher quality would be as equally distributed among the conditions as possible. However, because final teacher numbers were small, differences may be present in teacher quality among these three groups. Data regarding teacher characteristics were collected to assess any differences. Of the final number of intervention teachers participating in Year 1, 55% (six out of eleven) reported certification specific to grades nine through twelve. A total of 64% (seven out of eleven) met the districts' preference for having five or more years of teaching experience and 55% (six out of eleven) reported having master's degrees. It is important to note that teacher characteristics are not necessarily presumed to be indicators of teacher quality. In fact, the implementation levels were mixed for those teachers with higher levels of education, more years of experience, and certification in grades nine through twelve.

Finally, teachers who had not returned for Year 2 of the grant (four out of eleven) were evenly distributed across the districts and districts reported that the patterns of attrition did not differ from those normally observed. Districts also reported all teachers were placed based on their random assignment as planned in both years. If more than one position was open in Year 2, newly hired teachers were also randomly assigned.

[99] Xtreme Reading serves students reading at or above a fourth grade level as proposed.
[100] Control students were to be placed in these classes but the standard business-as-usual services were received in addition to ELA, which all participating students were to receive.

V. Evaluation of the Impacts of the Targeted Interventions

The Springfield and Chicopee School Districts implemented two targeted interventions for Striving Readers, READ 180 and Xtreme Reading in five high schools across the two districts.[101] The primary research question addressed by this study as required by the grant is: *Does participation in a reading intervention increase reading achievement?*

1. Does participation in READ 180 improve ninth graders' reading achievement as compared to the control group? If so, to what extent and what is the magnitude of the observed difference?
2. Does participation in Xtreme Reading improve ninth graders' reading achievement as compared to the control group? If so, to what extent and what is the magnitude of the observed difference?

To assess the effectiveness of the interventions, a randomized controlled trial (RCT) was employed. Eligible 9th grade students were assigned to one of three conditions: Control, READ 180, or Xtreme Reading.[102] Each of the treatment group impact estimates—for READ 180 and Xtreme Reading—were assessed in comparison to the control group. Because students were randomly assigned to intervention groups, students are the primary unit of analysis. However, the evaluation design includes the randomization of both students and teachers to one of the three conditions.[103]

[101] One additional high school in Springfield is not included in the grant and is not part of the study sample.
[102] Although these interventions were also implemented in the upper grades (10th, 11th, and 12th) as per the districts' request a control group was included only in 9th grade. Therefore, only 9th grade students were included in the impact analysis.
[103] Randomization of teachers was also conducted, which was possible because new teachers were hired with the agreement they would be placed at random in one of three positions: READ 180, Xtreme Reading, or Control (business as usual).

To answer the primary research question regarding the effectiveness of the interventions and to provide estimates of their "true" effects on reading achievement, average reading achievement scores of students in each of the two interventions were compared to the scores of students in control group classrooms, pooled across sites and study years.[104] Included in this report are the associated power estimates based on the numbers of students in the ninth grade cohorts.

Measures, Screening and Random Assignment

The primary outcome for the analysis of student impacts is the *Stanford Diagnostic Reading Test, 4th edition* (SDRT-4).[105] The SDRT-4 score comprises four key indicators of reading achievement: decoding (phonetic analysis), vocabulary, comprehension, and scanning.[106] This assessment was administered to all students school-wide, including struggling readers, by the districts in the spring of each year.

The *Scholastic Reading Inventory* (SRI) was used as the districts' screening tool as this assessment was already in use in some of their schools. The *Massachusetts Comprehensive Assessment System* (MCAS) English Language Arts test was used as the covariate in the analytic models to control for prior reading achievement level. The rationale for the inclusion of the MCAS as a covariate rather than the *Scholastic Reading Inventory* (SRI) is described in more detail in the analytic section.[107]

[104] Note that cohort in this instance is equivalent to year (e.g., Cohort 1 was treated in Year 1). Because students were randomly assigned to intervention groups, they are the primary unit of analysis.

[105] The SDRT-4 was also administered to participating struggling readers in the fall of each school year (2006-07, 2007-08) to further assess placement via the district screening process but later eliminated due to the burden on students and teachers.

[106] The SDRT-4 serves as both the outcome measure for the impact analysis as well as the screening measure for identifying struggling readers in grades 10-12 (which are not part of the RCT).

[107] The preliminary impact analyses conducted in the first year included the MCAS for seventh and eighth grade ELA separately to assess any potential impact use of the seventh grade MCAS would have. Although the eighth grade MCAS scores are used in these analyses the placement for those missing SRI scores in the first year was based on 7th grade MCAS scores. The correlation in the sample between the seventh and eighth grade MCAS scores is r =.56.

Exhibit 35 summarizes the data collection process, as well as the measures used for the estimation of student impacts.

Exhibit 35. Summary of measures and data collection schedule

Measure	Reliability	Schedule	Sample	Who Collects
Scholastic Reading Inventory (SRI)	Test-retest reliability[108] ranged from .78 to.97	Years 0-3: annually, spring *(baseline-screening)*	All 8th grade students	District
Massachusetts Comprehensive Assessment System (MCAS): English language arts (ELA)	Internal-consistency reliability[109], Cronbach's alpha .90	Years 0-3: annually, spring *(pretest covariate)* [a]	All 8th grade students	District
Stanford Diagnostic Reading Test-4 (SDRT-4)	Test-retest reliability Kuder-Richardson[110] .84 - .90 vocabulary; .91 to .94 comprehension; .88 to .93 scanning	Years 1-2: annually, fall *(placement assessed)* [b]	All 9th grade *striving readers*	District
		Years 1-4: annually, spring *(outcome)*	All 9th grade *(all students including striving readers)*	

[a] The SDRT-4 was to be administered annually in the fall but was eliminated in response to concerns regarding the testing burden on all parties (SR and non-SR combined).
[b] This test was administered to all students school-wide and used for non-RCT placement (students entering 10th, 11th, and 12th grade).

[108] Sources: http://research.renlearn.com/research/pdfs/57.pdf; http://www.proedinc.com/customer/default.aspx; Scholastic Professional Paper (March, 2006). Internal-consistency reliability was not reported.
[109] Source: http://www.doe.mass.edu/mcas/1998/techrpt_sum.pdf.
[110] Source: SDRT-4 Technical Manual, Harcourt, Inc.

Screening as Planned

All incoming 9th grade students identified as struggling readers based on the screening process were included in the pool for random assignment to interventions. The SRI has overlapping Lexile levels and, as a result, the range for identifying eligible incoming ninth grade struggling students had to be established (therefore, the 50th Normal Curve Equivalency or NCE was used as the benchmark).

Exhibit 36. SRI ranges from norms file: Unpublished data provided by Scholastic[111]

Student enrolled grade level (spring)	Reading level	Minimum SRI-Lexile score (50th NCE for 4th grade)	Maximum SRI-Lexile score (50th NCE for two grades Below)
8th	6th – 4th grade	680	855

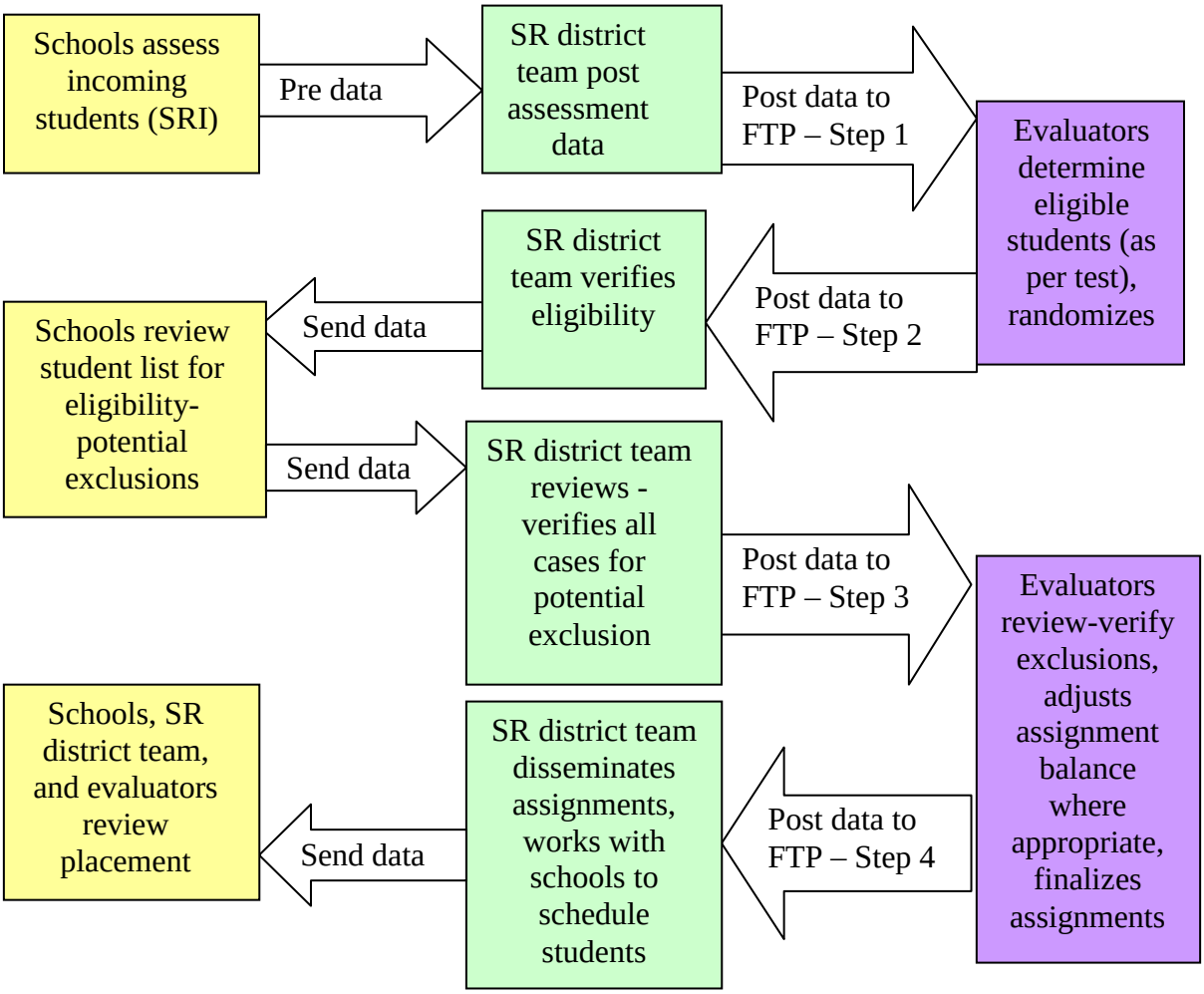
Districts established testing schedules and assessment protocols for the administration of screening. The SR district implementation team worked with the middle schools to screen the incoming 9th grade students in their final months of 8th grade to ensure they could be assessed for eligibility and scheduled as appropriate prior to the fall. The SR district implementation team worked with Scholastic to implement the SRI online so that it could be used for both assessing students at baseline and for monitoring progress in READ 180 over time. The districts provided the student test data which evaluators then used to randomly assign students.

Randomization Process as Planned

Approximately equal numbers of students were to be assigned to one of the three conditions. Randomization was conducted by the evaluator. The exhibit below represents the random assignment process as planned.

[111] Scholastic provided secondary data used to establishment this range or threshold.

Exhibit 37. Process and responsibilities for the final randomization using SRI test scores (ninth grade screening test) [112]



[112] School and district responsibilities are one in the same but referred to here as "school" responsibilities. FTP is the file-transfer protocol site established by the evaluator to maintain data confidentiality as per data sharing agreements.

Pre-randomization blocking of students (by special education and English language learner status) was employed where numbers permitted, to ensure the similarity of students across groups on observable characteristics relevant to the outcome and to increase the precision of impact estimates.[113] Sample size estimates did not exceed the districts' ability to serve; therefore all those students screened and eligible were to be included in the pool to be randomly assigned.[114]

Following the receipt of SRI scores, evaluators randomly assigned students to one of the targeted interventions or the control group. This process occurred over approximately a one-week period, if complete data were provided including grade, school, and state identification number as well as other data used for assignment within strata. Once randomized, students were excluded from the study if they met any of the following criteria: (1) their Individual Education Plans (IEPs) explicitly specified a different form of reading support; (2) they lacked the necessary English language or comprehension skills; (3) their parents formally refused participation in the interventions;[115] (4) they were enrolled off-campus in a "twilight school," an evening program without a Striving Readers' program, or in an "early college high school," a college preparation program; [116] (5) they had high grade histories and MCAS scores that were at least proficient; or (6) they were deemed "inactive" by the districts, meaning that the district was not able to determine whether they were enrolled in any of the schools.

[113] The constraint placed on the range of struggling readers to be identified left little opportunity to block on levels of screening status (Xtreme Reading serves only those students reading at a fourth grade level or higher).

[114] Students who were reading below a fourth grade reading level would not participate in the study but were to receive the supports and interventions normally provided by the district (i.e., business as usual). Special education students whose Individual Education Plans (IEPs) stipulate that they receive services different from the interventions were excluded from the study. Students enrolling in schools after the fall verification period (mid-October) would not participate in the study that school year.

[115] Parents with questions about student placement spoke to the coordinators in either district, and then discussed concerns with the vice principals or principals. If, after an explanation of the study and placement parents still requested the student be removed, they were asked to provide a letter stating their request to not have their child participate and the student was removed from the intervention class.

[116] Off-campus enrollment was the case only in SPS.

Students who transferred between any of the five high schools across the two districts were *not* excluded from the study, but were scheduled into the same condition to which they had been previously randomly assigned in their original school. In each district, expected rates of exclusions including general attrition were unknown and therefore estimated by the evaluator (not by the districts). Finally, district and school staff members were to review the assignments and discuss any concerns with evaluators as well as potential exclusions. For details regarding exclusions by cohort, district, etc., refer to Appendix I.

Research protocols and requirements were established whenever possible in collaboration with the SR district implementation team. The district maintained responsibility for communicating with their staff regarding all Striving Readers activities. However, the SR district implementation team worked with evaluators to distribute information about the research study, schedule information sessions at staff meetings, and hold question-and-answer sessions about the study at each of the schools. Refer to Appendix E for examples of protocols and information provided to district staff.

Teacher Recruitment and Assignment

During the 2006-07 school year teachers within each of the five participating schools were randomly assigned to teach struggling readers in READ 180, Xtreme Reading, or business-as-usual English Language Arts classes (i.e., the 9th grade control group). Although teachers were also randomly assigned to either intervention in grades 10 through12, districts were unwilling to include a control group in these grades. In the 2007-08 school year, open teaching positions were filled as needed and teachers were randomly assigned if more than two positions were open at a time.

Evaluators randomly assigned teachers to the three conditions based on information districts provided regarding their backgrounds. Wherever possible, evaluators stratified assignment based on number of years of teacher experience (two or fewer years teaching) so that teachers new to the profession were assigned and equally distributed across the three conditions. When known, within-district experience was considered as well.

After random assignment, intervention teachers participated in READ 180 and Xtreme Reading training beginning in August 2006.

Per the districts' final implementation plan, the school districts intended to hire a total of 40 "reading literacy teachers." However, final district estimates included 30 newly hired teachers and 10 teachers already employed by the districts, due to the recruitment challenges in Springfield (including an absence of teacher contracts and layoffs the first year).

Although each district decided it would hire teachers individually as their own district employees, they had agreed to use the same job description to ensure that any qualified teacher would be considered qualified in both districts. The job description per the district implementation plan listed preferences for new teacher hires, including: (1) certification in English or reading or in the process of attaining either, (2) five years of experience in teaching English or reading, (3) some experience in the use of technology for teaching, and (4) availability to attend summer professional development training. In addition, teachers hired for the positions had to agree *as a condition of their employment* to be randomly assigned to one of three conditions: Control, READ 180, or Xtreme Reading. Teachers could not request or choose which condition they were to teach as per hiring requirements.

The Striving Readers Chief Implementation Officer submitted job postings to district human resources staff and distributed copies to school principals for use in interviewing and hiring over the summer. Recruitment venues included local school district job fairs, internal job announcement posting sites, and local newspapers. Striving Readers district staff confirmed that principals provided the "Letter of Teacher Expectations" developed by the Striving Reader Chief Communications Officer to all teachers prior to their official hire date. The letter listed job-specific requirements (e.g., collaborating with district staff for data collection, attending professional development as required, adhering to the intervention specifications if assigned to one of the treatment groups, and collecting and reporting student data on reading achievement). In some cases, principals from Springfield and Chicopee actively involved the grant coordinator for Springfield and the SR district implementation team leads in the interviewing and hiring processes.

Final Sample

There were several factors which influenced the final sample size of teachers and students. Many included barriers related to existing school schedules, context, etc. One complicating factor in the first and second years of the grant was the conversion of one of the five participating high schools to Pilot School status. Because this school had not made adequate yearly progress for several consecutive years, the Massachusetts State Department of Education mandated that it become a Commonwealth Pilot School in 2007-08, the second year of the Striving Readers implementation, as explained earlier in this report. This conversion entailed restructuring the school to become five schools-within-a-school and the notification and planning for this process began in the first year of the Striving Readers implementation.

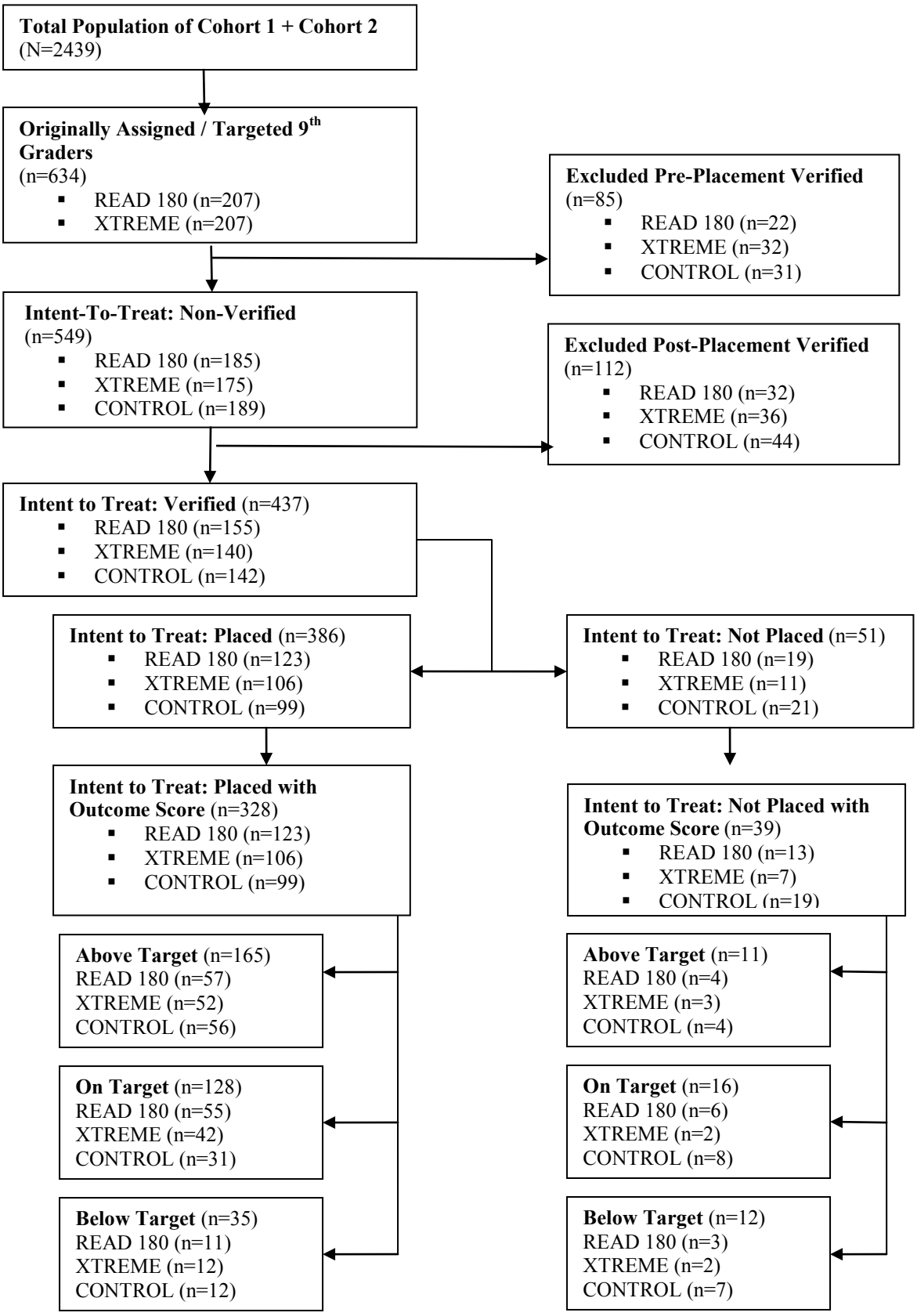
The actual numbers of teachers hired and assigned were initially based on the numbers of striving readers identified in the screening process and ultimately on the final numbers of those striving readers returning in the fall. Fewer teachers than anticipated were included in the study given that: (1) screening complications resulted in fewer total students assessed; and (2) initial estimates of the qualified population included *all* students with reading abilities two levels below grade down to a first-grade level (rather than a 4th grade level, which SIM developers later specified was the lower threshold of effectiveness for the Xtreme Reading program).

In Year 2 of the grant, developers clarified that teachers were not formally required to sign a contract specifying the requirements of their position for this grant. So, contracts were essentially standard district teaching contracts. Given the reduction in qualifying students, many of the SR teachers were used to teach other district and school courses, some of which were intervention classes in the upper-grades as well as other general education courses.

Student Screening and Random Assignment

Currently three cohorts of 9th grade students from the 2006-07, 2007-08, and 2008-09 school years have participated in the RCT and the final cohort will participate in the 2009-10 school year. All cohorts will be combined for the final analysis of targeted intervention impacts.

Exhibit 38. Screening and assignment samples



During the initial Year 1 random assignment process, 90 students (21%) of the total number of students in Springfield were identified as striving readers based on their 7th grade MCAS ELA test scores because the spring SRI test scores were not available. Evaluators obtained reading scores from the MCAS as an alternative way to screen those students who were not screened with the SRI to identify struggling readers. However, none of these students remained in the intent-to-treat or ITT sample (none enrolled in the fall as per district records).

Verification and Exclusions

Students were excluded from the study as per the criteria described pre-placement. However, the majority of students were excluded based on valid criteria post-placement (39% as compared to 61%, respectively), occurring in the fall after students were already placed in their classes. Refer to Exhibit 39. Post-placement attrition comprised 40% of the total exclusions overall. There were 51 students excluded post-placement (21%) who had been eligible for placement. For specific information regarding verification of reportedly excluded cases, refer to Appendix I.

Exhibit 39. Final numbers of the excluded students - by school

Assignment		**Cohort 1 and Cohort 2 Total**					
		CCHS	CHS	Commerce	Putnam	SciTech	TOTAL
Valid Exclusions Pre Enrollment	9th	6	13	21	24	12	76
Valid Exclusions Post Enrollment	9th	5	4	54	22	36	121
Not Placed	9th	1	3	12	21	14	51
	Total	**12**	**20**	**87**	**67**	**62**	**248**

Cohort 1, 2, and 3 (Years 1, 2, and 3, respectively) samples are included in subsequent tables presenting power estimates. The verified placement percentages were based on the numbers of those expected for placement (i.e., those in the ITT group).

Intent-to-Treat

The following exhibit presents the final number of students in the Intent-to-Treat (or ITT) condition for Years 1 and 2.

Exhibit 40. Final numbers of the Intent-to-Treat randomly assigned students - by school

Assignment		Cohort 1 and Cohort 2 Total					
		CCHS	CHS	Commerce	Putnam	SciTech	TOTAL
Control	9th	25	16	21	33	26	121
READ 180	9th	23	22	28	32	31	136
Xtreme Reading	9th	24	16	21	36	32	129
Not Placed	9th	1	3	11	22	14	51
	Total	**73**	**57**	**81**	**123**	**103**	**437**

Note that 51 students, approximately 12% of the ITT group, had initially been reported inactive by the districts but were actually in attendance at least 75% of the time, based on both rosters and district attendance.

The following exhibits present final projected sample information. For more information regarding cohort samples in Years 3 and 4 and projected total impact sample, refer to Appendix I.

Exhibit 41. Numbers of randomly assigned and placed cohort students (cohort = year)

Assignment Grade 9	Total				
	Included in Current Analysis [117]		*Verified to Date*	*Expected*	
	Cohort 1	Cohort 2	Cohort 3	Cohort 4	TOTALS
Control	67	54	59	56	236
READ 180	72	66	58	57	253
Xtreme Reading	70	57	62	57	246
Verified (actually placed % of ITT)	209 (75%)	177 (67%)	179 (70%)	170 (68%)	735 (70%)
Eligible-originally assigned then expected placement (ITT)	334 (279)	300 (264)	303 (254)	300 (250)	1237 (1047)

Note: The initial screening results indicated the number eligible or originally assigned (those performing at least two grade levels below their current grade level but no lower than a fourth grade level).

The following exhibit presents the numbers of placed students by treatment condition for the anticipated sample.

Exhibit 42. Number of the randomly assigned and placed cohort students

Assignment Grade 9	Total	
	Cohorts 1 - 3	Cohorts 1 - 4
Control	180	236
READ 180	196	253
Xtreme Reading	189	246
Verified (actually placed % of ITT)	565 (71%)	735 (70%)
Eligible-originally assigned then expected placement (ITT)	937 (797)	1237 (1047)

[117] In Year 1, 285 were reportedly placed. However, based on rosters provided in the second year, there were only 279 confirmed to be enrolled.

Screening and Placement Barriers

In the first two years, major barriers to obtaining student information in order to verify placement included the following:

- Initial screening complications related to technology requirements, process, and communication regarding test importance;
- Unknown school attrition and exclusions rates;
- Incomplete placement and verification information (data were entered by the SR district implementation team rather than district-generated. Given the timeline in Year 1, repeated revisions to student status often occurred but were often unavoidable.); and
- Student-level data from the district was not available and/or reported until the following school year (e.g., enrollment, attendance rates, etc., which follow state and federal reporting schedules).

Communication challenges regarding grant requirements and implementation plans were influential barriers. There were many challenges in the first year regarding implementation of the program and the evaluation; protocols and processes were established while the implementation design and subsequent evaluation specifications were being finalized and testing and random assignment were taking place.

Approximately half of the interviewed teachers across districts reported that they believed some of their students were "misplaced" into the program and were too advanced in terms of literacy skills to be included in a targeted intervention classroom. This was described as more problematic in Year 1 and data provide evidence to support their assertions. Students did score higher in the first year as compared to the second, yet there were no group differences between cohort (i.e., scores within year were similar across treatment and control groups). Cohort differences were not anticipated given eligible students represented a distinct and narrow sample of those in the population and there were no major shifts in the student population in general.

Other teachers noted students had told them the tests were not taken "seriously," they hadn't realized at the time the test was "important," and therefore they did not perform at their best. A process for verification of student data was established to determine whether or not a student's test score was an accurate representation of his or her reading ability.[118] Districts were to review each case referred or questioned by the school staff (after assignment lists were received for prior MCAS scores as well as grades in ELA) as a way in which to verify reliability of the screening.

As reported by the SR district team, they became aware over time that teachers were not privy to the differences in testing systems and the procedures which had been put in place for schools to review each student assignment. In subsequent testing, according to the SR district team, measures were taken to communicate to administrators, teachers, test proctors, students' family members, and to students themselves, that all tests for SR were absolutely critical and would determine placement in and out of literacy interventions. Students were specifically advised that tests should be approached with seriousness and with the intention of scoring as highly as possible. In order to address any misunderstanding regarding the screening process in Year 2 of the grant, the Striving Readers Work Group reported developing materials including proctor scripts and written procedures (e.g., a timeline, procedural steps, and accompanying checklists) to assist middle schools with administering high-quality screening of the students in eighth grade (using the SRI as per district plans). Through these methods, the SR district team sought to ensure that any misunderstandings regarding screening were addressed in Year 2. Interview data in Year 2 suggests an increase in participating teacher understanding of study requirements including those related to screening and placement.[119]

[118] In the second year, evaluators received student level data to pre-identify eligibility based on ELL and SPED status but these were reported by the district team in the first year.

[119] Teacher knowledge of the plans, the importance of adhering to study guidelines, and the significance of having consistency across districts reportedly increased in Year 2 as well.

Finally, the planning time allotted for the implementation prior to the initial screening round (the SRI testing in spring of 2006) and the start-up of Year 1 was not optimal. Specifically, difficulties arose as a result of the complexities associated with implementing the grant, such as: (1) an abbreviated timeframe between grant award notification and the startup of the grant; (2) adjustments to the study design (as mandated but resisted by districts) to include a "true" control group; (3) initially insufficient technology set-up to screen students in a timely manner; and (4) overcoming an end-of-the-school year timeframe that challenged the planning, communication, and buy-in process. The result was that the initial screening (using the SRI) of students in select schools occurred at lower than optimal percentages. In addition, changes in key district staff created further difficulties in initial and "make-up" screenings and slowed communication between districts and the evaluator regarding assignments and the planned assignment review process at the school level.

Power to Detect Effects

Minimum detectable effect size (MDES) estimates have been computed and are used to determine whether or not the study design provides sufficient power to detect an impact if one exists for either intervention. The MDES indicates how small an effect the intervention can have on students reading achievement and still be detected (Orr, 1999). Effect sizes are reported on a scale of 0 to 1, and the higher the score, the greater the magnitude of the treatment effect (Cohen, 1998; Lipsey, 1990). The framework used to assess the magnitude of effect sizes was Cohen's (1988): .20 as small, .50 as moderate, and .80 or above as large (as cited in Bloom et al., 2005).[120] More recent research provides other empirical benchmarks for evaluating effect sizes related to education-focused interventions (Hill, Bloom, Rebeck Black, & Lipsey, 2007; Vernez & Zimmer, 2007).

[120] Observed effects characterized by both Lipsey's theory-based framework are remarkably similar to Cohen's empirically-based framework: 0.15-0.20 as small, 0.45-0.50 as medium or moderate, and 0.80-0.90 or above as large.

Whenever possible, normative expectations for change, gaps in effect among subgroups, and effects reported for like-interventions were examined for information ensuring power estimates were more accurate.[121] A meta-analysis conducted by Borman et al. (2003) analyzing achievement effect size data from more than 145 studies of various forms of comprehensive school reform, observed effect sizes of .07 (25^{th} percentile) to .25 (75^{th} percentile). Recent research on high school intervention models related to literacy noted effect sizes on achievement ranging from .17 to .20. Analysis of National Education Longitudinal Study data yielded effect sizes of .17 unadjusted and .08 adjusted for covariates (Bloom et al., 2005). Methodologically sound studies of various educational technology initiatives (bearing a partial similarity to READ 180) considered in a recent review, present similar effects (Murphy, et al., 2001; Waxman, Connel, & Gray, 2002).

Given these findings, Vernez and Zimmer (2007) recommend interpreting effect sizes from data related to educational interventions aimed at positively impacting student achievement levels as follows: 0.05-0.10 as small, 0.15 as medium or moderate, and 0.25 as large.[122]

Power for Analysis

Initial power estimates were based on a two-level framework and the planned assignment of teachers/classes. However, the number of teachers was fewer than anticipated and resulted in only one teacher per condition, per school—effectively rendering equal to school-level in these analyses (which is insufficient for multilevel modeling using classroom as the cluster).

[121] Other considerations include the properties of the achievement measure and whether effect sizes are calculated using levels of achievement or growth. Achievement growth is steeper and more pronounced in the earlier years and therefore effect sizes are generally higher in the earlier school years as compared to the later school years. Hill et al. (2007) provide data indicating reading growth effect size by grade was roughly .94 or more in early elementary years; .57 to .25 in the later elementary through middle school years, respectively; and .20 to .03 in the early to later high school years.

[122] "This interpretation is supported by Lipsey and Wilson's (1993) review of meta-analyses across psychological, educational, and behavioral outcomes, which concluded that effect sizes of 0.10 to 0.20 should not be seen as trivial" (Vernez, & Zimmer, 2007).

Each of the two intervention groups of students (Xtreme Reading and READ 180) will be compared to the control group of students; therefore, there were two pair-wise comparisons of striving readers in the ninth grade cohorts pooled across years 1, 2, 3, and 4 as now proposed. Refer to the following exhibit. These were tested using the full sample in both regression models (using effect-dummy coding) and in the ANCOVA models.

Exhibit 43. Pooled ninth grade student sample for pair-wise comparison

Condition Total	Xtreme Reading	READ 180	Control	Pair-wise Total [123]
Cohort 1 – 2				
First Pair	128	—	128	256
Second Pair	—	128	128	256
Cohort 1 – 3				
First Pair	188	—	188	376
Second Pair	—	188	188	376
Cohort 1 – 4				
First Pair	245	—	245	490
Second Pair	—	245	245	490

Note: The actual combined sample of 565 is divided into the three conditions (n ~ 188 for Control, READ 180, and Xtreme Reading). The estimated Cohort 4 combined sample of 735 is divided into the three conditions (n ~ 245 for Control, READ 180, and Xtreme Reading).

Current MDES calculations were calculated for a single-level trial as developed under Optimal Design (Raudenbush & Liu, 2001; Raudenbush, Spybrook, Liu, & Congdon, 2004). In Year 1, estimates of the correlation coefficients between pretest scores, or prior achievement scores, and post-test scores at various levels were made in the absence of the availability of actual data (Raudenbush, et al., 2004; Bloom, 2004).

[123] The final numbers of randomly assigned and placed ninth grade students are distributed approximately equally across the five high schools and among the three study groups (READ 180, Xtreme Reading, and Control).

Specifications for the power estimates in Year 2 remained the desired 80% power to detect an effect with two-tailed tests of significance (at the .05 significance level). However, the pre- and post-test achievement score correlation was unknown at the time of initial calculation.[124] The following exhibit presents the power estimates for the pooled cohort samples, including the MDES with the pretest covariate.

Exhibit 44. MDES for pair-wise comparisons: by N of students and covariate r (.80 power, 5% significance level, two-tailed test)

Number of Students		Minimum Detectable Effect Size (σ)	
		By Covariate Correlation	
		No covariate	*r = .52**
2 Cohorts	*N = 256 per contrast*	.35	.31
3 Cohorts	*N = 376 per contrast*	.29	.25
4 Cohorts	*N = 490 per contrast*	.25	.21

The MDES estimate based on the projected student numbers is .35 for the two-cohort study. Including the MCAS ELA prior achievement score as a covariate with an r^2 of .27 lowers the MDES estimate to .31 for the two-cohort study. The MDES estimate is .29 for the three-cohort study and .25 for the four-cohort study; with the pretest covariate they are .25 and .21, respectively. Results approximate those presented in current research scenarios estimating sample size for randomized trials though many of those estimates include higher pretest covariate correlations (Bloom, 2005).

[124] This correlation between pretest covariate (MCAS) and post-test outcome (SDRT-4) was .53 for the combined sample.

Blocking was conducted for student assignment by school and grade as well as by disability and ELL status, which should increase the precision of estimates (Raudenbush, Martinez, & Spybrook, 2005). Although blocking by screening level was initially proposed, it was not ultimately pursued due to the restricted reading-level threshold (two levels below grade down to a fourth grade level) imposed by the Xtreme Reading developers. This threshold yielded a smaller pool of striving readers than originally anticipated.

Statistical Analysis

The analysis is designed to estimate the impact of the two interventions separately by comparing the achievement scores of each treatment group on average to that of the control group. Using reading scores from standardized assessments taken in the spring of the ninth grade year, student performance in reading for each of the two treatment groups will be compared with the control group.[125] As described previously, given projected and actual power estimates, third (2008-09 school year) and fourth (2009-10 school year) cohorts with control groups will be added, thus yielding a larger than originally planned sample to include in the final impact analyses.

Although there was random assignment of students (and teachers),[126] students remain clustered within schools and, if clustering is not accounted for, the standard errors could be miss-specified and overestimate treatment effects. However, given the limited teacher sample (i.e., for ninth grade only, there are not multiple teachers per condition), the multilevel models fit using HLM were not ideal or stable given the very small numbers of clusters or schools (Raudenbush & Bryk, 2002).

[125] As per district request, after one year, students in the ninth grade control groups are randomly assigned to one of the two interventions for tenth grade if they are not yet reading at or above grade level.
[126] Recall that students are the primary unit of analysis

Multilevel models were fit in Year 1 and though they eventually converged, it was deemed more appropriate given the primary unit of assignment to utilize a fixed-effects approach using OLS regression; four indicator variables were entered for the five high schools in the final model.[127] Because the school-level sample size will remain small (n=5) throughout the grant, the fixed-effect model will continue to be appropriate to use.[128] These analyses were conducted but the decision was then made to use ANCOVA for ease of interpretation.

Pretest Scores

There were three primary issues with the pretest or screening placement assessment used: (1) a lower than anticipated number of ninth grade students being screened with the SRI in two schools in particular, resulting in the need for dual screening assessments (the SRI and the MCAS ELA);[129] (2) the absence of a strong relationship with the outcome scores; and (3) higher than anticipated standard errors.

The number of students missing the SRI but who had prior MCAS scores (n=25) was low in the final analytic sample for ninth grade; only 17 of the 25 had SDRT-4 outcome scores. Therefore, initial plans to conduct an assessment of those originally placed via the MCAS (n=90) were modified and a common metric was not required. However, although significant correlations were observed between tests, the SRI demonstrated lower correlations with the MCAS (r =.18) grade eight as well as with the SDRT-4 (r =.25), which were much lower than reported correlations by Scholastic between the SDRT-4 and the SRI.

[127] The HLM analysis conducted yielded an intraclass correlation of .19; that is, the amount of variance in the reading scores to be predicted between groups i.e., schools is 19%, while the variance to be predicted at the individual level is 81%. This intraclass correlation is consistent with similar research on school effects and the predominance in cross-sectional data of the individual characteristics (Bloom et al., 2005; Raudenbush & Bryk, 2002).

[128] These decisions were made in collaboration with the evaluator's TA provider, Abt Associates. For one of the smaller schools, there were fewer than 10 students per treatment condition severely limiting the power in an HLM structure thus the planned analysis will combine data across cohorts. As the student sample size increases each year, power will be reevaluated and the appropriateness of fitting multi-level models will be assessed (TA communication).

[129] This fact presented a challenge in that the interpretability of a beta parameter depends on the associated covariate having a common metric. If a covariate does not have a common metric, the beta parameter may capture differences in the metric itself rather than the underlying construct, which in this case is reading ability (measured by the assessment).

Given that no prior reading achievement data were available before the administration of the SRI as is strongly recommended by the developer to incorporate prior known status, there were concerns regarding the higher than anticipated standard errors. Based on discussions with the technical assistance provider and the fact that the spring SDRT-4 scores were more highly correlated with the 8th grade MCAS scores (r =.53) than with the SRI, it was decided that these MCAS scores would be used as the pretest covariate in the impact models.

Analytic Model

The dependent variable (outcome) used to estimate the impact of the targeted intervention on students reading achievement is the Stanford Diagnostic Reading Test version 4 (SDRT-4). The outcome, the SDRT-4 standardized-scale scores, used in the experimental impact models is continuous.

The independent variables included in the analysis of impacts were:

- the treatment indicators (Xtreme Reading, READ 180, and Control);
- MCAS eighth grade pretest scores; and
- student-level demographic characteristics (covariates):
    - special education status;
    - English language learner status;
    - race/ethnicity;
    - socio-economic status as measured by free and reduced lunch status;
    - gender; and
    - age (age over time of test included based on the difference in promotion/retention policies between districts).

Average achievement across schools was tested. Refer to the following exhibit.

Exhibit 45. Variables included or tested for inclusion (covariates in impact models)

Measurement Construct	Variable	Level	Coding/Range [a]	Comments
An indicator of whether the data come from the first or the second 9th grade cohort	Cohort_0_1	Student	=1 if data obtained from a student from Cohort 2 =0 if Cohort 1	To be tested for significance. Note: Addition of Cohort 3 (variable not to be included if not significant)*
Gender	Female	Student	=1 if female =0 if male	Based on district CCD coded data regarding status (time invariant)
English Language Learner status	ELL	Student	=1 if student is classified as an English Language Learner =0 else	Based on district CCD coded data regarding status
Free and Reduced Lunch status eligibility/ classification – a proxy for socio economic status	Free_Lunch	Student	=1 if student classified as free or reduced price lunch =0 else	Based on district CCD coded data regarding status
Special Education status	SPED	Student	=1 if student is classified as special education =0 else	Based on district CCD coded data regarding status
Race/ethnicity	Minority	Student	=1 if student is classified as a minority (district codes) =0 else	Based on district CCD coded data regarding status (race/ethnicity state codes will be changing as per districts over time)
Age	Age_over15	Student	=1 if student is over 15 years of age (district codes) =0 else	Based on district data reported for student date of birth (calculation based on April-May time of screening)
School	Sch1, Sch2, Sch3, Sch4	School	=1 if school # =0 else	Based on district data (dummy coding)
MCAS_pre	escaleds8_fnl	Student	Continuous (RANGE 200-280)	MCAS ELA performance scores from Grade 8 (pretest covariate)

Note: A variable to indicate imputed missing SRI scores (if a regression-based model was used) would have been included if the SRI was to be included as the pretest. If Z-scores were to be used, a dichotomous variable indicating type of pretest would have been included. However, as explained, the SRI is no longer used as the pretest covariate.

Model Specifications

Analyses were designed to answer the research question, "Does participation in READ 180 improve ninth graders' reading achievement relative to that of a control group?" using students as the primary unit of analysis. The model for this cross-sectional analysis of the impact of the targeted intervention is specified with fixed effects for schools. In other words, the overall impact of each targeted intervention is estimated as a treatment effect averaged across schools.

The model includes the baseline/pretest score as a covariate (now the MCAS ELA scores from grade eight). Model covariates assessed for inclusion in the final model included student-level characteristics coded as dummy variables: race/ethnicity, free and reduced lunch status, special education status, English language learner status, minority status, gender, as well as a variable indicating whether the student was over age for their grade at pre-assignment. Cohort and school differences were assessed in the models. ANCOVA models were fit for these analyses allowing the effects of participation in the interventions to be separately assessed in the same model. The dependent variable, reading achievement, was measured on a continuous scale (i.e., using SDRT-4 scaled scores).[130] Treatment effect size estimates were calculated.

Decision Rules for Variables

Covariates were initially all included in the models as a block. Empirical evidence was used to keep or remove covariates and the final model was specified based on the inclusion criteria established for the removal of the covariates ($p<.20$). Baseline equivalence testing and correlations among coefficients (to identify possible multicollinearity) were conducted using regression. Post-hoc model assessments were also conducted to assess multicollinearity and to determine the potential influences of outliers (using Cook's distance).

[130] For binary outcome measures (reading at grade level/reading below grade level or GLE) an analogous logistic regression model will be used. These models will be specified in future analyses.

Missing data were assessed for each variable and did not exceed 20% missing (TA communication).[131] Of those with MCAS pretest scores, 79% had SDRT-4 post-test scores. Missing data for this sample was less than 20% the threshold set for imputation: Control = 26 (17.9%), READ 180 = 17 (11.1%), and Xtreme Reading = 27 (19.4%). As a result, models presented included the non-imputed sample of pre-and post-test scores. The assumption that missing data were missing at random (MAR) was made, given there were not differences between groups (treatment and control) in rates of "missingness."

Exhibits 46 and 47 present descriptive information about the sample by district and treatment group, respectively. Characteristics are presented for the combined cohorts and the ITT sample. Refer to Appendix I for a presentation of these data within cohort.

Exhibit 46. Student sample characteristics by district (n=437)

Characteristics	District		Total (freq/mean)
	Chicopee	Springfield	
Race/Ethnicity (%)			
White	92.3	43.0	252 / 57.7
Black	6.2	49.8	161 / 36.8
Asian	0.8	1.0	4 / 0.9
American Indian	0.0	1.0	3 / 0.7
Other*	0.8	5.2	17 / 3.9
Female Gender (%)	46.2	60.3	245 / 56.1
Special Education Status (%)	18.5	11.7	60 / 13.7
English Language Learner Status (%)	1.5	7.2	24 / 5.5
Free and Reduced Lunch Status (%)	56.2	86.3	338 / 77.3
Attendance (mean)	162.4	154.4[a]	156.8[b]
MCAS Score (mean)	233.8[c]	228.6[d]	230.1[e]
SRI Score (mean)	782.8[f]	779.6[g]	780.6[h]
Sample size (*n*)	130	307	437

Note: Sample size for calculation [a] 305 [b] 435 [c] 112; [d] 298; [e]410; [f]129; [g]280 and [h]409

*Other includes combinations of White, Black, Asian, American Indian, Native Hawaiian, and Hispanic.

[131] The necessity of outcome imputation was assessed as planned for the combined cohort analysis. Results did not differ, and correlations were stronger for the non-imputed pretest covariate and the outcome score. Dependent upon the status of missing data in future cohorts, imputation (MI) methods using HLM and NORM, a program developed to provide estimates for missing data (Raudenbush et al., 2004; Shafer, 1997), will be potentially employed.

As illustrated in the exhibit above, aggregate student characteristics differ between districts. These reflect differences between districts on variables reported for their populations. Across all students included in the preliminary analysis sample and assessed at baseline, more than half were non-minority students with the majority in Springfield (over 56%). In addition, Springfield has significantly higher ($p<.05$) numbers of females than Chicopee (60% versus 46%, respectively). There were significant differences when the threshold for significance is considered higher ($p<.10$) among Common Core Data (CCD) collected by district including special education students, those classified as English language learners, and those with free and reduced lunch status. In this student sample, 86% in Springfield as compared to 56% in Chicopee qualify for free or reduced-price lunch, a proxy used to represent student socio-economic status ($p<.001$). These variables represent indicators of students potentially at-risk for poor reading achievement. Note that the sample sizes between the districts differ (the balance is 30% versus 70%) which may influence the significance of the differences observed. Mean attendance level, defined as the number of days present since October 1 of the total, did not differ significantly between districts.

Exhibit 47. Student sample characteristics by treatment (n=437)

Characteristics	Intervention			
	Control	**READ 180**	**Xtreme Reading**	**Total (FreqMean)**
Race/Ethnicity (%)				
White	59.9	54.8	58.6	252 / 57.7
Black	36.6	**40.6**	32.9	161 / 36.8
Asian	0.0	0.6	0.7	4 / 0.9
American Indian	0.0	0.0	1.4	3 / 0.7
Other*	2.8	3.2	7.1	17 / 3.9
Female Gender (%)	52.1	**61.9**	53.6	245 / 56.1
Special Education Status (%)	12.0	12.3	17.1	60 / 13.7
English Language Learner Status (%)	4.2	5.8	6.4	24 / 5.5
Free and Reduced Lunch Status (%)	76.8	73.5	**82.1**	338 / 77.3
Attendance (mean)	157.9	156.8 [a]	155.7[b]	156.8[c]
MCAS Score (mean)	230.4 [d]	230.3[e]	229.6[f]	230.1[g]
SRI Score (mean)	778.1 [d]	780.7[h]	783.2[i]	780.6[j]
Sample size (*n*)	142	155	140	437

Note: Sample size for calculation [a] 154 [b] 139 [c] 435; [d] 135; [e] 146 ; [f]129; [g]410; [h]144; [i]130; and [j]409

*Other includes combinations of White, Black, Asian, American Indian, Native Hawaiian, and Hispanic.

The following exhibits present the data for the pre-post ITT analytic sample (n=347). Patterns remain the same, with some differences, though non-significant at the p<.05 level between groups. These differences did not appear systematic but may be the result of the smaller sample sizes within each group.

Exhibit 48. Student sample characteristics by district: pre-and post-test sample (n=347)

Characteristics	District		Total (freq/mean)
	Chicopee	Springfield	
Race/Ethnicity (%)			
White	93.3	41.3	198 / 57.1
Black	4.8	53.7	135 / 38.9
Asian	0.0	0.4	1/ 0.3
American Indian	0.0	0.4	1 / 0.3
Other*	1.0	4.5	12 / 3.5
Female Gender (%)	47.6	62.0	200 / 57.6
Special Education Status (%)	18.1	12.0	48 / 13.8
English Language Learner Status (%)	1.0	5.0	13 / 3.7
Free and Reduced Lunch Status (%)	55.2	87.2	269 / 77.5
Attendance (mean)	167.3	162.3[a]	163.8[b]
MCAS Score (mean)	234.0	229.4	230.8
SRI Score (mean)	782.5[c]	778.7[d]	779.9[e]
Sample size (*n*)	105	242	347

Note: Sample size for calculation [a] 241; [b] 346; [c] 226; [d] 330[e]

*Other includes combinations of White, Black, Asian, American Indian, Native Hawaiian, and Hispanic

Students in both districts were similar on the SRI reading achievement assessment screen and the MCAS as would be expected if the same group of targeted students was being identified, though Chicopee students in this sample score higher on average. Analysis results indicate that, on average, the random assignment process was generally effective, given sample sizes, in creating equivalent groups based on the variables measured and those used in stratification (SPED, ELL).

Exhibit 49. Student sample characteristics by treatment: pre- and post test sample (n=347)

Characteristics	Intervention			
	Control	**READ 180**	**Xtreme Reading**	**Total (FreqMean)**
Race/Ethnicity (%)				
White	61.1	52.3	58.5	198 / 57.1
Black	37.2	43.8	34.9	135 / 38.9
Asian	0.0	0.0	0.9	1 / 0.3
American Indian	0.0	0.0	0.9	1 / 0.3
Other*	1.8	3.9	4.7	12 / 3.5
Female Gender (%)	51.3	61.7	59.4	200 / 57.6
Special Education Status (%)	11.5	12.5	17.9	48 / 13.8
English Language Learner Status (%)	3.5	3.9	3.8	13 / 3.7
Free and Reduced Lunch Status (%)	78.8	73.4	81.1	269 / 77.5
Attendance (mean)	164.9	162.4 [a]	164.4	163.8[b]
MCAS Score (mean)	231.0	230.7	230.7	230.8
SRI Score (mean)	777.9[c]	777.0[d]	785.4[e]	779.9[f]
Sample size (*n*)	113	128	106	347

Note: Sample size for calculation [a] 127 [b] 346 [c] 109; [d] 119; [e] 102 ; [f]330

*Other includes combinations of White, Black, Asian, American Indian, Native Hawaiian, and Hispanic

Students pretest and baseline covariate scores (SRI and MCAS) were similar across groups, though the SRI scores were higher for the Xtreme Reading group in comparison to the others.

V.A. Preliminary Impacts on Students

The impacts presented briefly in this section are preliminary. In future reports, cohorts will be combined to establish the power needed to conduct the initially proposed analyses. In the interim, fully specified models have been fit as requested for review and discussion among evaluators at the annual Striving Readers meeting.

In the following exhibit, the outcome or SDRT-4 scores are presented as Grade Level Equivalencies (GLE). This figure presents the results for all students screened and placed based on their reading performance. The percentage of students attaining grade level reading expectations was 12% for both Control and Xtreme Reading students and 14% for READ 180 students. The percentage of students reading *below* Striving Readers eligibility was high (as measured by the outcome, a different assessment from the screening assessment). However, the percentage of students with reading skills below the fourth grade level was higher in the control group at 42% as compared to READ 180 at 35% and Xtreme Reading at 38%. It is unclear why so many students reading scores were below the targeted level. This finding is potentially indicative of some or all of the following factors: difficulties in placement related to differences between cohorts, variations between SDRT-4 and SRI ranges, and reliability of SRI scores.

Despite increases in average GLE reading scores, there were no statistically significant differences observed between the treatment groups and the control group.

Exhibit 50. Impact on reading achievement grade-level equivalencies (GLEs)

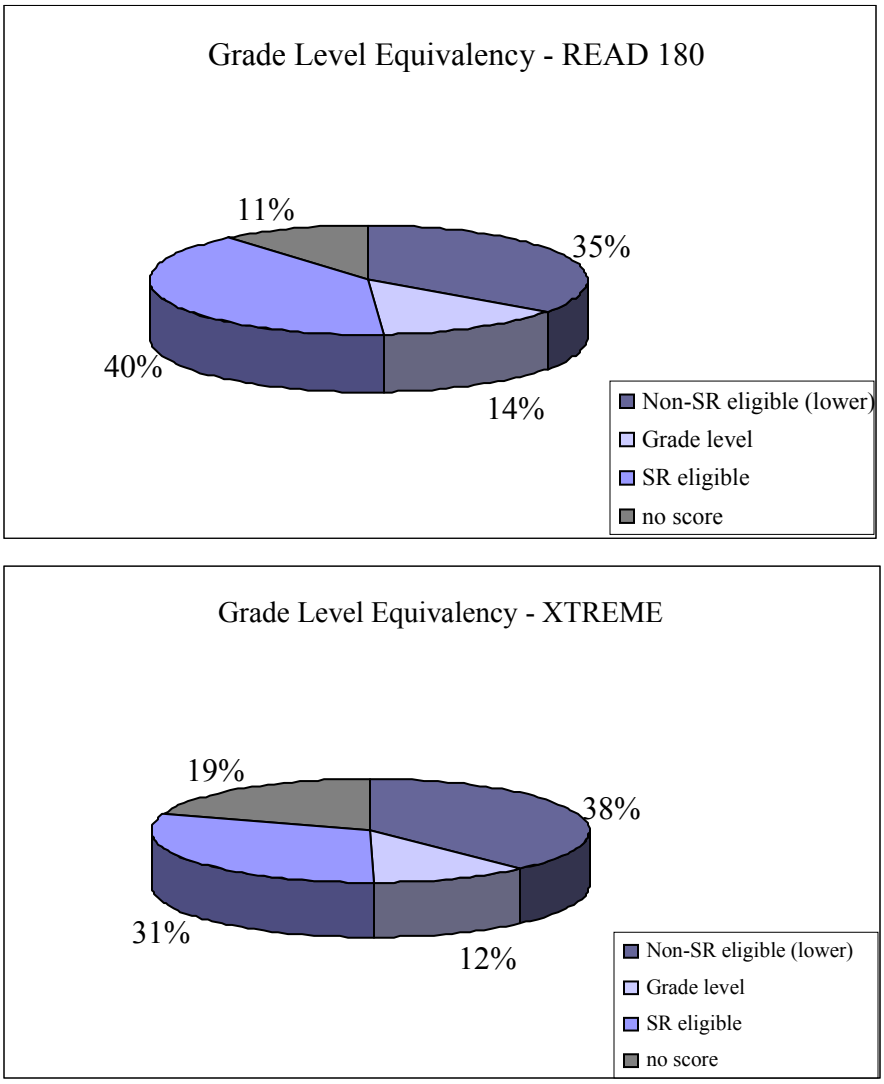


Exhibit 51 illustrates the impact of each intervention on student reading achievement. The exhibit presents both unadjusted and regression adjusted impact estimates. The results are presented in concert with effect sizes. The final model includes school by treatment interaction terms and their inclusion may result in an over-specification of the models. The interactions were based on the effect-coded school variable (with interpretation of results for the fifth school). Refer to Appendix I for results from the earlier models that were fit and for those without the school variables, etc.

Exhibit 51. Impact of intervention on student reading achievement (SDRT-4)

	Unadjusted Means			ANCOVA-adjusted Means		
	Control	**Treatment**		**Control**	**Treatment**	
		READ 180	**Xtreme Reading**		**READ 180**	**Xtreme Reading**
Reading Achievement Mean	669.11	671.02	671.54	661.94	664.78	666.09
SD	25.74	27.80	26.93	45.99	47.36	43.64
Estimated Impact	--	1.91	2.43	--	2.84	4.15
Effect Size [a]	--	.07	.09	--	.11	.16
P-value	--	.62	.98	--	.31	.16
Number of Students [b]	114	128	105	114	128	105
Number of Schools =5						

[a] Effect sizes were calculated (Glasses) for unadjusted means using the control group standard deviation.
[b] Sample for the regression-adjusted model was dictated by the numbers with both pre- and posttests (n=347 of the ITT n=437).

As the table illustrates, there were no observed (significant) effects of the interventions as compared to the control group. As stated in the technical assistance provider memo:

> *In the ideal (i.e., when random assignment works perfectly), the difference between these two means would be the unbiased estimate of program impact. However, all sites are planning to use covariates to adjust the model to help guard against bias that may have been introduced because random assignment did not work perfectly. The regression adjusted means and impact estimate will reflect these adjustments.*

The adjusted means were lower than the unadjusted means. There was no observed mean difference between treatment and control significant at the p<.05 level. Though statistically non-significant, the minimal effect of both treatment groups was increased based on these estimates. Effect size estimates included were Glasses' Δ (Abt communication; Rosenthal, 1994).[132] This formula is described as:

$$\text{SES} = \frac{(\bar{y}_{treatment} - \bar{y}_{control})}{\text{standard deviation of control group}}$$

The analytic model presented included only covariates significant at the p<.20 level. The model then was re-fit to include only significant covariate-predictors given fewer degrees of freedom due to smaller sample sizes. Outcomes for those models with the full complement of predictors assessed were extremely similar to those models fit with only significant covariates.

Finally, using criteria outlined by What Works Clearinghouse (WWC) for assessing the rigor of designs and analysis, baseline or pretest scores were assessed to identify pre-treatment differences among the groups. No significant differences were observed among the groups. Pretest scores were not observed for the three groups (two treatments and one control) to be over a .50 standard deviation difference. In addition, the numbers of "actual" exclusions were examined to identify differential attrition between groups (i.e., these exclusions would have been noted at the time of screening and assignment review but were not available to evaluators until early fall, particularly in Springfield). No differences in attrition estimates among treatment groups were greater than 20%.

[132] Quoted from Abt TA guidance: Rosenthal (1994) refers to Option 1 as "Cohen's *d*" or "Hedges's *g*", and refers to Option 2 as "Glass's Δ" and states "the pooled S – that is, the 1 computed from both groups – tends to provide a better estimate in the long run, of the population standard deviation (Hedges & Olkin, 1986, p.79). However, when the S's based on the two different conditions differ greatly from each other, choosing the control group S as the standardizing quantity is a very reasonable alternative. That is because it is always possible that the experimental treatment itself has made the S of the experimental group too large or too small relative to the control group (p. 232)."

Exhibit 52. Impact of intervention on student reading achievement (SDRT-4): final model

Tests of Between-Subjects Effects

Dependent Variable: SDRT4 (READING ACHIEVEMENT)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	91245.336[a]	11	8295.031	17.641	.000	.367
Intercept	74487.441	1	74487.441	158.415	.000	.321
Treatment	975.359	2	487.679	1.037	.356	.006
ELL	1228.846	1	1228.846	2.613	.107	.008
SPED	1623.059	1	1623.059	3.452	.064	.010
Minority	3313.450	1	3313.450	7.047	.008	.021
School	9274.016	4	2318.504	4.931	.001	.056
MCAS ELA Grade 8	38258.083	1	38258.083	81.365	.000	.195
Cohort Year	7121.530	1	7121.530	15.146	.000	.043
Error	157518.428	335	470.204			
Total	1.563E8	347				
Corrected Total	248763.764	346				

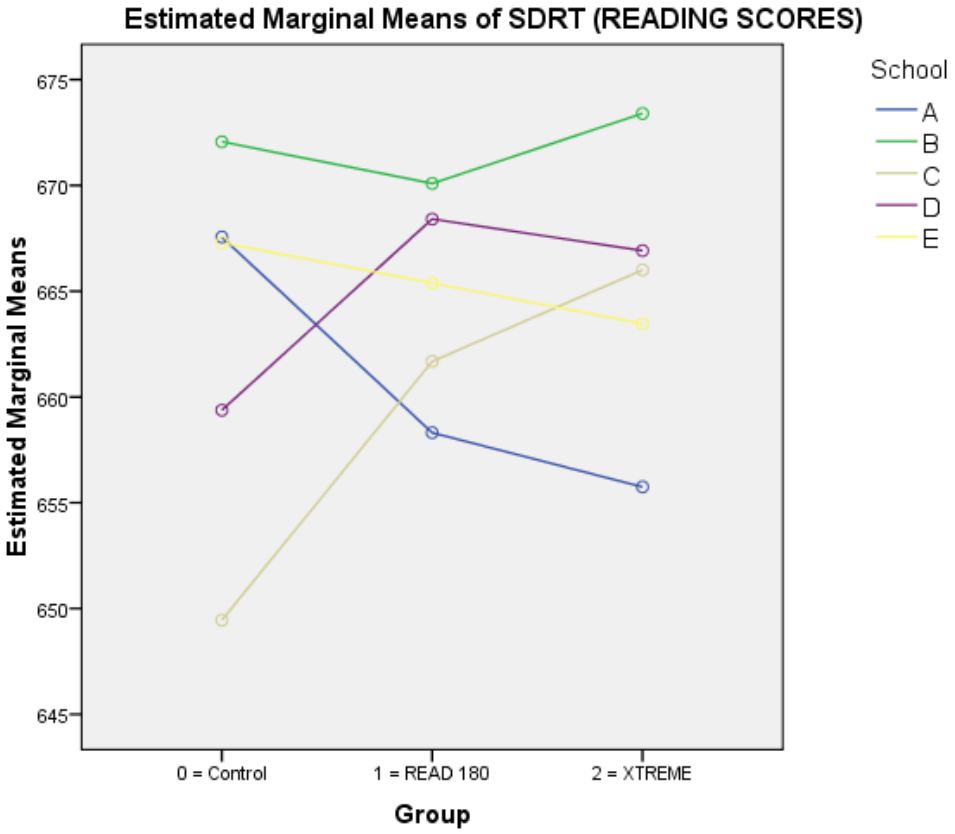
a. R Squared = .367 (Adjusted R Squared = .346)

An assessment of treatment and school interactions indicated no significance at the $p<.05$ level (nor were those tested via regression models with pretest covariate). There were cohort differences overall with higher outcome scores in the first cohort as compared to the second, however this result was consistent across groups. That is, outcome scores in Cohort 1 were higher than those in Cohort 2; they were higher for the treatment and control groups alike. Given the challenges in first year implementation of screening and placement, the decrease may reflect more accuracy in the screening process. This hypothesis has some support from the interview data in the second year. These data suggested that students took the tests more seriously; that teachers and test administrators communicated the importance of the tests to students more clearly; and that the systems were in place to verify data.

Although scores in Cohort 2 were higher overall than Cohort 1, there were two schools for which outcome scores remained high. As illustrated in Exhibit 53, one school scored unusually high in this sample (especially given the comparison to the population of the school), and the patterns for the remaining four were split in accordance with the direction of the effect. However, two schools had slightly larger increases for the treatment groups as compared to the control than the other two schools which scores decreased relative to the control, which accounts for the overall higher means in the combined sample. It is important to note that the two schools with very high control group scores were also the schools with high overall outcome scores in both cohorts.

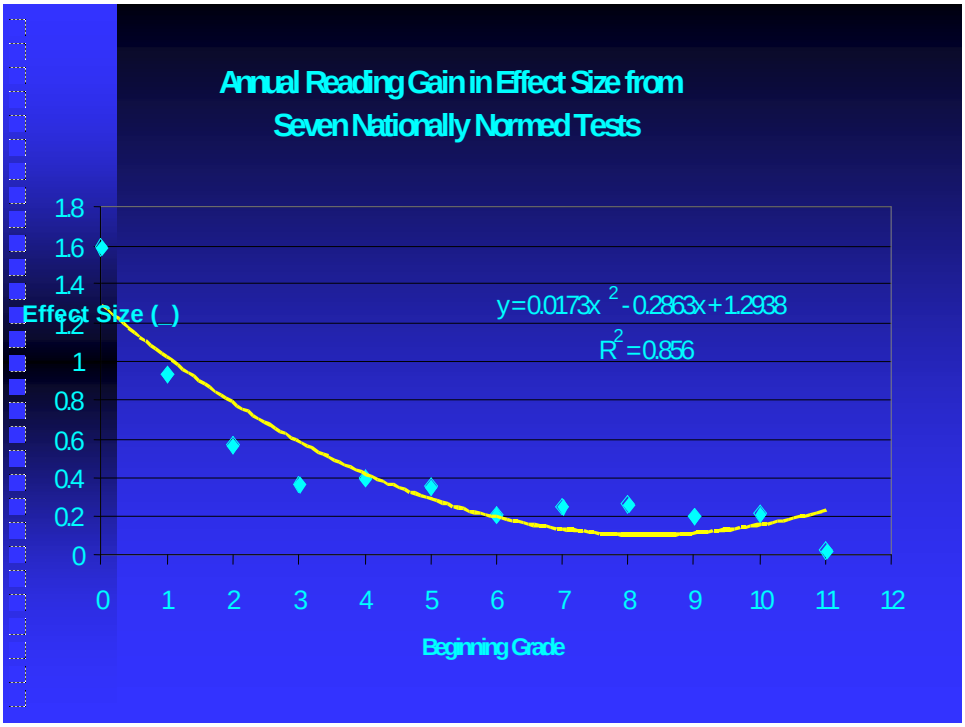
Finally, as the exhibit below also illustrates, the performance of interventions *within* each school was similar relative to the control group. That is, student scores generally increased for both groups of intervention participants or decreased for both groups of intervention participants (with the exception of the single school with inexplicably high ratings overall as compared to the population at this school).

Exhibit 53. Plot of results for treatment groups by school



As illustrated in Exhibit 54, when achievement gains are assessed across grade level, effect sizes decrease in the upper grades. Therefore, striving readers in the high schools would generally be expected to gain less than those in the lower grades simply as a result of the trajectory of student growth or development of reading skills.

Exhibit 54. Empirical benchmarks for achievement gains in reading



Source: Bloom, Hill, Rebeck Black, & Lipsey, 2006

V.B. Impacts and Implementation

Targeted Impacts and Implementation

The goal of the targeted implementation study was to inform the interpretation of impact findings by *describing* the context in which the interventions were implemented. More specifically, implementation levels were established to characterize the context and its complexity and, as a result, to provide a gauge by which to judge any observed effects relative to the context. Therefore, the following analysis was purely exploratory and not intended to predict the impact of the interventions.[133]

Describing the implementation context in relationship to observed impact involved several steps. The first step was to combine classroom implementation ratings across two years in order for this information to more accurately represent the context of the combined cohort data assessed in the impact study.[134] Overall ratings were calculated by adding ratings across years and dividing by the total number of possible items to be rated, thereby weighting the scores (tables are provided in Appendix I and J).[135] The second step involved summarizing the implementation levels to represent both study years combined as had been done for each individual year with the following four levels: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

[133] The hypothesis that higher levels of implementation would be related to higher levels of observed impact was not empirically tested; analyses were purely illustrative. As described in the Enhanced Reading Opportunities Study, such analyses: "…are not able to establish causal links between these aspects of implementation and variation in program impacts across sites, because school characteristics and other implementation factors may confound the association between…impacts and the implementation factors included in the exploratory analysis" (Corrin, et al., 2008).

[134] Classroom implementation was used to describe context for this purpose. Input levels were previously discussed as influences on classroom implementation context in concert with other non-intervention factors (e.g., school).

[135] It is important to remember these data were collected in snapshots and by definition represent only a picture of implementation at that precise point-in-time.

The third step involved examining the implementation and impact results together for each intervention to identify emergent patterns. This examination was also conducted across interventions to illuminate any overall patterns which may have emerged across both interventions.

Finally, a discussion of this analysis is provided at the conclusion of this section.

READ 180

The comparison of implementation and impact results for READ 180 is included in the exhibit below. This exhibit illustrates that, in schools where classroom implementation levels were observed to be the highest, average reading scores of READ 180 students were higher relative to students in the control group.[136]

Exhibit 55. READ 180 classroom implementation level and impacts: Years 1 and 2

	Implementation		**Impact** [137]
School	**Rating**	**Level**	**Change Relative to Control**
A	42%	Low	Negative
B	55%	Moderate	None (slightly lower)
C	**83%**	**Adequate**	**Positive**
D	**58%**	**Adequate**	**Positive**
E	42%	Low	None (slightly lower)

Note: Averages were calculated weighted by the total number of items across years. Implementation levels: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

[136] Note the two schools with "low" levels of implementation had rates fairly close to the threshold for a "moderate" implementation level.

[137] The title "impact" is not used to note statistical significance but is a description of the increase or decrease of student scores, on average. The power to detect a meaningful effect was not yet sufficient and this exploratory analysis was conducted only to examine data for patterns between score changes and implementation ratings.

That is, READ 180 student scores were higher at post-test, controlling for pretest scores and other student characteristics; though this difference was not statistically significant.

Xtreme Reading

The comparison of implementation and impact results for the Xtreme intervention is included in the table below. The following exhibit illustrates that, for half of those schools with higher classroom implementation levels (all moderate in this case), average reading scores of Xtreme students were higher relative to students in the control group.

Exhibit 56. Xtreme Reading classroom implementation level and impacts: Years 1 and 2

	Implementation		**Impact** [138]
School	**Rating**	**Level**	**Change Relative to Control**
A	33%	Low	Negative
B	55%	Moderate	None (slightly higher)
C	**55%**	**Moderate**	**Positive**
D	**56%**	**Moderate**	**Positive**
E	64%	Moderate	None (slightly lower)

Note: Averages were calculated weighted by the total number of items across years. Implementation levels: No evidence (0 - 24%), Low (25 - 49%), Moderate (50 - 74%), and Adequate (75 - 100%).

That is, Xtreme student scores were higher at post-test, controlling for pretest scores and other student characteristics; though this difference was not statistically significant.

[138] The title "impact" is not used to note statistical significance but is a description of the increase or decrease of student scores, on average. The power to detect a meaningful effect was not yet sufficient and this exploratory analysis was conducted only to examine data for patterns between scores changes and implementation.

Preliminary Intervention Patterns

In this exploration of implementation and impact, a pattern emerged across interventions. For both interventions, student reading scores increased as compared to control student reading scores in two of the five schools; the same schools in which implementation levels were higher in general. Although this pattern was apparent, the increases were not statistically significant on average. Schools without increased student reading scores had more variable implementation levels, especially in the case of Xtreme. Therefore, the positive pattern across interventions (i.e., higher intervention levels and higher reading scores) may be related to the context, school in this case, in which the models were implemented.[139] Note this pattern holds more strongly for READ 180 than for Xtreme Reading.

Discussion

Despite the many complications related to implementation, particularly in Year 1 of the study, a pattern of higher targeted implementation levels and higher overall student reading scores was observed. Although neither intervention's observed difference in impact scores as compared to the control group was significant, this may be attributed to the fact that overall sample size was small. Two more cohorts will participate in this effectiveness trial in the next year which will increase and potentially double the current sample size to one at which statistically significant differences may be discernible.[140] The descriptive results discussed here may *foreshadow the potential* for detecting meaningful intervention effects under conditions such as an increased sample size and increased classroom implementation levels.

[139] There were at least three teachers in every school (one per group), and in the first year teacher was equal to the school for each group. In the second year, the majority of teachers were new and therefore, in the combined analyses, many schools have more than one teacher representing results. Note this remains a small sample of teachers/classes representing "school."

[140] Significance was not tested within individual schools given the sample size and a lack of power to detect within school effects.

Although analyses were conducted for both years combined, implementation levels and impact results varied by year, which itself has implications and at a minimum requires caution when interpreting the findings. It is important to note that these cautions should be exercised for both interventions, as there were differences in implementation between years for both Xtreme and READ 180. In Year 2, overall classroom implementation ratings increased over those in Year 1 for READ 180 (47% and 63%, respectively), and decreased for Xtreme (60% and 49%, respectively). In addition, the ratings in Year 2 for READ 180 teachers who replaced those from Year 1 were higher[141] (all but one teacher was replaced). However, the opposite pattern was observed for Xtreme teachers who replaced Year 1 teachers; these teacher's ratings were lower than those from the previous year and similar to the ratings of the two remaining Year 1 teachers. For Xtreme implementation ratings, there was less observed variability in Year 2 as compared to Year 1 relative to that in READ 180.

Year 2 results were potentially confounded by the number of new teachers and their respective backgrounds as well as the increased specificity of observation tools (use of two observations versus one for ratings and the inclusions of additional subcomponents in the classroom observation). It is difficult to disentangle the possible influences of newly hired teachers from the changes in scoring specificity; collectively these two influences appeared to be positive for READ 180 but negative for Xtreme. The addition of two more cohorts from which implementation data will be collected should facilitate the ability to separate out the unique influence of each these factors.

[141] One of the replacement teachers in Year 2 had the same level of implementation as the prior Year 1 teacher (adequate). However, the Year 1 teacher's rating was slightly higher overall at 100%.

VI. Whole School Intervention: Implementation Study Design

Research Questions and Methods

Similar to the approach for examining implementation of the targeted interventions, implementation research questions and data collection activities were developed for the SIM-CERT whole school intervention. Exhibit 57 includes specific whole school implementation research questions based on the program models and their intended activities, methods, objectives and ultimate outcome goals.

Across the areas of implementation, data collection served multiple purposes: (1) to document and assess fidelity of implementation; (2) to determine the level of program implementation; (3) to document variation in program implementation; and (4) to examine variation in program implementation as a potential influence on observed outcomes.[142] Data were also collected to assess the presence of relevant contextual factors for SIM-CERT teachers and additional district and school staff responsible for implementing the school-wide intervention.

Exhibit 58 lists the data sources used to answer each of the evaluation questions in Year 2 given the status of first year implementation. Scoring is described in more detail in the following section in which implementation levels are presented.

[142] Outcomes for the whole school intervention will be included in subsequent reports.

Exhibit 57. Specific implementation research questions: Whole school intervention

What was the level of implementation and variability of professional development /support for teachers/administrators/literacy coaches?

*Professional development – **initial** training from developers:* [143]

Teachers

- What proportion of teachers received/participated at *different levels* in the initial professional development?*
- What proportion of teachers received/participated in the initial professional development at *an adequate level*?**

Administrators[144]

- What proportion of administrators received/participated at *different levels* in the professional development?*
- What proportion of administrators received/participated in the initial professional development at *an adequate level*?**

*Professional development – **ongoing** training from developers:*

Teachers

- What proportion of teachers received *different levels* of ongoing training?*
- What proportion of teachers received *an adequate level* of ongoing training?**

Literacy coaches

- What proportion of literacy coaches received different levels of ongoing training?*
- What proportion of literacy coaches received an adequate level of ongoing training?**

*Professional development – **ongoing** mentoring from literacy coaches:* [145]

- What proportion of teachers received *different levels* of ongoing mentoring by coaches?*
- What proportion of teachers received *an adequate level* of ongoing mentoring by coaches?**

What was the level of implementation and variability of classroom instruction?

- What proportion of teachers implemented the classroom model (frequency of SIM-CERT routine use) *at different levels* of implementation?*
- What proportion of teachers implemented the classroom model as specified by the developers at *an adequate level* of implementation?**

Note: In both exhibits, one asterisk (*) is used to specify cases in which components of the intervention is examined by level of implementation (e.g., majority of the time, most of the time, some of the time, almost never). Two asterisks (**) are used to specify cases in which both the appropriate level of implementation and the proportion evidencing this level of implementation were used to examine intervention implementation.[146]

[143] Initial in this context for teachers and administrators is defined as training that took place in the planned summer professional development (PD) period prior to each cohorts' initial year implementation. This PD is considered to be the foundation for program implementation.

[144] Districts reported these trained took place, but data were not reported at the individual level for inclusion here.

[145] On-site CERT-trained literacy coaches provided ongoing mentoring (classroom observations and demonstrations, problem-solving, feedback, etc.) for teaching staff. Data related to ongoing mentoring will be provided and included in subsequent reports.

[146] Exhibits were developed by Abt Associates, the technical assistance provider to Striving Readers evaluators.

Exhibit 58. Research questions data sources: Whole school implementation study

Research Questions*	Measures/Data Sources**						
	Surveys/ Interviews				District Records/ Records Review		
	Teacher surveys	Teacher focus groups	District-school administrative staff	Literacy coach	Professional development attendance	Curricula, district-provided developer materials	Class rosters (scheduling)
What was the level of implementation and variability of professional development/support for teachers/administrators?							
Professional development/support (PD) for teachers, administrators, and literacy coaches (initial, ongoing workshops and on-site mentoring)							
Proportion of teachers, coaches, and administrators receiving **different levels** of *initial* professional development*	√	√			√	√	
Proportion of teachers, coaches, and administrators receiving **adequate level** of *initial* professional development**	√	√			√	√	
Proportion of teachers and literacy coaches receiving **different levels** of *ongoing* training*	√	√		√	√	√	
Proportion of teachers and literacy coaches receiving **an adequate level** of *ongoing* training**	√	√		√	√	√	
Proportion of teachers receiving **different levels** of *on-site professional development via coaches* *	√	√		√			

Research Questions*	Measures/Data Sources**						
	Surveys/ Interviews				District Records/ Records Review		
	Teacher surveys	Teacher focus groups	District-school administrative staff	Literacy coach	Professional development attendance	Curricula, district-provided developer materials	Class rosters (scheduling)
Proportion of teachers receiving **an adequate level** of *on-site professional development via coaches* **	√	√		√			
What was the level of implementation and variability of classroom instruction?							
Proportion of teachers who implemented the classroom model at different **levels***	√	√		√			
Proportion of teachers who implemented the classroom model at **an adequate level** **	√	√		√			

Note: Refer to the footnotes on the prior exhibit regarding what data were and were not provided and included in analyses.

Whole School Implementation Data Collected

The implementation data collected via each method identified in the previous table is detailed below. Measures used are provided in the appendices. To evaluate the dosage and intensity of the implementation of the whole school intervention SIM-CERT, data were collected from four primary sources: (1) surveys administrated to teaching staff at the five participating high schools; (2) interviews conducted with literacy coaches, school, and district administrators; (3) focus groups conducted with randomly selected SIM-CERT-trained teachers at each of the five participating high schools; and (4) district and developer records. All data collection activities were conducted in Years 1 and 2 with exception of the focus groups, which were conducted only in Year 2 to provide additional context for implementation over time.

Teacher surveys

The SIM-CERT teacher survey (see Appendix F) was designed to elicit participant professional development session experiences, knowledge of instructional routines, and patterns of the use of routines among teachers. The survey was administered to all teaching staff, regardless of whether they had participated in SIM-CERT training, and will be administered throughout the period of the grant, to assess teacher participation in professional development and the prevalence of SIM-CERT routine knowledge and use over time. In Year 1, the survey was provided to district staff to distribute at planned meetings in April 2007 and was collected in sealed envelopes to maintain the confidentiality of the participants.[147] A total of 452 teachers returned completed paper surveys in year 1, yielding a 66% completion rate across all schools. In Year 2, given more time to coordinate with the districts to obtain the necessary email addresses of teachers, the survey was administered online. A survey link, embedded within a personalized email to every teacher, was sent in April 2008 and remained open for one month.

[147] Participants could choose not to record their name to maintain their anonymity. The option to record their name would allow the linkage of their responses with district provided professional development attendance data if they choose to do so.

Evaluators sent two follow-up emails to remind teachers to complete the survey. A total of 472 school personnel returned completed surveys, yielding a 77% completion rate[148] across all schools. The Striving Readers program team sought to improve the response rate in both years by providing scripts to principals for making an announcement about the survey either over the intercom, during faculty meetings or via email. This script explained the purpose of the survey, emphasized its confidentiality, and scheduled time for its completion. Additional efforts to increase the response rate included the provision of incentives; teachers who completed the survey participated in a lottery drawing (random selection) for $25 gift cards.

Because the whole school analysis plan requires data be collected over four years of implementation, the survey results contained in this report provide an abbreviated picture of fidelity to the model as planned. In future reports, survey results for the entire teaching staff (SIM-CERT and non-SIM-CERT teachers) will allow evaluators to identify the prevalence of SIM-CERT routine knowledge and use over time.

Literacy Coach and Administrator Interviews

Literacy coach and administrator interviews (refer to Appendix G) were designed to gather more in-depth contextual information regarding SIM-CERT planning, implementation and monitoring from the administrator perspective. In Years 1 and 2, individual interviews were conducted with SIM-CERT literacy coaches and administration staff within each school, including principals, assistant principals, and English Language Arts department chairs. Other key staff members were identified for interviews by the district team who could provide information about SIM-CERT implementation, including Special Education directors, guidance counselors, Instructional Leadership Specialists, and other district administrators. These data were collected May to June. In Year 1 and Year 2, interviews were conducted as a part of the planned evaluation activities for district staff.

[148] Response rate calculated as follows: the number of respondents/(the number of emails sent out – emails bounced back). The numerator does not include respondents that opted out of the survey. The rate was calculated prior to the removal of certain cases in the analysis.

SIM-CERT Teacher Focus Groups

Group interviews with SIM-CERT teachers were conducted in April 2008 at each of the high schools and were designed to gather more in-depth contextual information about CERT implementation from the perspective of participating teachers (see Appendix H). Topics addressed in focus group sessions include: (1) experiences with SIM-CERT training and coaching, (2) use of SIM-CERT in the classroom and (3) factors that support or impede use of SIM-CERT routines. Focus group participants, a total of 32 CERT-trained teachers, were randomly selected by the evaluators from each of the five participating schools and ranged in cohort as well as subject areas and grade levels taught. Districts had budgeted for incentives and payment for focus group participant time during Year 2.

District records (professional development attendance and other materials)

District records and developer records (e.g., professional development agendas) were collected over time to document the implementation of the model. These data were used to determine implementation scores of the professional development model for the second year and, in retrospect, the first year of the intervention. Survey data were used to calculate scores for fidelity to the classroom model for the second year of the intervention only. The calculation of scores for the first year report was not possible given incomplete attendance rosters (i.e., not provided per person) and professional development planned activities and details regarding classroom model expectations (developers continued to refine these). Therefore, the specifications were not defined until after the first year of the grant.[149] Initially districts provided all information but later in the second year of the grant this information was requested of developers. Again, SIM-CERT specifications evolved over time especially in the first year of the grant for several reasons including the fact that developers and districts determined scheduling and topics for training and these differed between districts. In addition, the developer made adjustments to training including materials in response to district and teacher and administrator feedback as well as based on their own philosophy of ongoing development and refinement.

[149] Professional development and classroom model requirements were being developed in Year 2 as well.

Whole School Implementation Teachers

Characteristics of SIM-CERT teachers and survey respondents (Year 1 and Year 2)

According to district documents, the first-year SIM-CERT-trained cohort consisted of 110 teachers and 5 literacy coaches for a total of 115 of the original 125 proposed. Of these 110 teachers, 90 responded in the SIM-CERT survey that they had received training provided by the developers.[150] Among survey respondents (n=452) drawn from the entire population of teaching staff at the five participating high schools, 45 teachers in Chicopee and 45 teachers in Springfield reported receiving SIM-CERT training during the first year of SIM-CERT implementation.

As shown by Year 1 survey data, there were slight deviations from the selection criteria established by the district team in that the teachers who received SIM-CERT training consisted not only of English, social studies, and science teachers, but also teachers of mathematics and other subjects. The survey respondents, specifically those who stated that they were trained in SIM-CERT during the 2006-07 academic year, reported teaching courses in the following areas: ELA and/or Xtreme Reading (n=21), history/social science (n=21), science (n=19), math (n=18), and other courses (n=18). Of the 90 respondents, 86 identified what courses they taught.[151]

District documents indicate that during the second year of implementation, Cohort 1 consisted of 101 teachers and Cohort 2 with 130 teachers for a total of 231 of the original 250 proposed. For both Cohort 1 and Cohort 2, Chicopee either met or slightly missed model requirements as planned (Cohort 1 = 48 of 50; Cohort 2 = 50 of 50). In Springfield, the number of teachers trained in SIM-CERT was less than expected for Cohort 1 but exceeded requirements for Cohort 2 (Cohort 1 = 53 of 75; Cohort 2 = 80 of 75).[152] The decreased number of CERT-trained Springfield teachers in Cohort 1 is likely due to attrition and the lower number of teachers initially included in Cohort 1 during the 2006-2007 school year.

[150] The number of teachers who reported receiving SIM-CERT training (n=90) is lower than the number reportedly trained (n=110). District documents indicated that by the end of the 2006-2007 school year, 110 teachers had attended all professional development workshops and had received SIM-CERT routine training.
[151] Note that some respondents indicated that they were teaching courses in more than one of the areas listed above.
[152] These categories are not mutually exclusive.

Teacher attrition and other barriers to teacher inclusion in SIM-CERT will be discussed in the implications section. Ninety-five of the 123 SIM-CERT-trained Cohort 2 teachers completed the Year 2 survey. Additionally, 55 of the 101 teachers from the first-year cohort also completed the Year 2 survey. Thus, a combined total of 150 SIM-CERT trained teachers responded to the Year 2 survey.

For both cohorts in the two districts, the majority of the Year 2 survey respondents indicated that they were certified at the professional level (59% and 64% SPS, 67% and 79% CPS, each cohort respectively). Survey respondents who were trained in SIM-CERT reported teaching courses in a variety of subject areas. The exhibit below reveals that across districts, the trained teachers in Cohort 2 who responded to the survey taught a broader range of subjects than their counterparts in Cohort 1 as per the original selection criteria as planned. However, as shown in survey data from Year 1 as well, both districts included math and other subject area teachers in the original cohort, a slight change from the PD model as planned. Evaluators found variation between districts in terms of the process of teacher recruitment for CERT although selection criteria had been established. For example, while teachers in one district tended to be recruited on a volunteer basis, teachers from the second district tended to be "assigned" to the CERT whole-school intervention. In Year 3, evaluators will aim to obtain more information on this topic.

Exhibit 59. Subjects taught by district and cohort

	Cohort 1		Cohort 2	
Subject	**Chicopee**	**Springfield**	**Chicopee**	**Springfield**
English (n=30)	8 (24%)	2 (9%)	7 (21%)	13 (21%)
Science (n=26)	7 (21%)	5 (23%)	7 (21%)	7 (12%)
Math (n=25)	9 (27%)	5 (23%)	4 (12%)	7 (12%)
Social Studies (n=27)	9 (27%)	8 (36%)	3 (9%)	7 (12%)
ESL (n=15)	0	0	4 (12%)	11 (18%)
Special Education (n=7)	0	0	5 (15%)	2 (3%)
Other (n=10)	0	0	4 (12%)	6 (10%)
Multiple Subjects (n=9)	0	1 (5%)	0	8 (13%)
No Subject Listed (n=1)	0	1 (5%)	0	0
Total (n=150)	33	22	34	61

Note: Numbers reflect respondents who taught *only* that subject except for the 'multiple subjects' row.

Cross-checking teacher-reported training attendance with district PD lists from the first and second years of the program revealed a number of discrepancies. In both cohorts, there were discrepancies between teachers' cohort membership as reported by the district and cohort membership as reported by the teachers themselves based on survey responses indicating their attendance of CERT training. Discrepancies could be the result of the lag time between survey and training or misunderstanding among survey respondents as to what constitutes CERT training (e.g. mandatory SIM-developer provided training versus school-embedded coaching support).

VII. Whole School Intervention: Level of Implementation and Implications

Professional Development Ratings

As with the two targeted interventions, ratings were created to establish the level of adequacy of implementation of the professional development and classroom models. Adequacy is defined as the implementation of intervention components as specified by the developers and the districts in their final plans and as depicted in the whole school literacy intervention logic model. It is assumed that model components are specified at the level necessary to establish change in content literacy via student use of the various SIM-CERT routines.

Two subcomponents were included in the overall rating of the level and adequacy of planned SIM-provided professional development: (1) receipt of the *initial training workshops before the first year* of each cohort's implementation of the intervention, (2) receipt of *ongoing training workshops within the academic school year* that built upon the planned initial training provided. According to model specifications, two initial and two ongoing training sessions were required for teachers during their first year of teaching SIM-CERT.[153] During the second year of teaching SIM-CERT strategies, two additional ongoing training sessions were recommended.[154]

[153] Information was provided by both developers and district staff regarding requirements of the PD model. Each SIM-developer-provided professional development session included six hours of training in the implementation of SIM-CERT routines in the classroom.

[154] In addition to the model-specified rating criteria, the districts requested that the evaluators provide an alternative framework for determining whether the SIM-CERT training model has been implemented as planned. The alternative framework considered the percentage of teachers in Cohort 1 who attended five total training sessions over two academic school years, without distinguishing the timeframe in which professional development was delivered. However, data were not provided by one district on ongoing training sessions scheduled for June 2008 and corresponding make-up sessions in August 2008. With the addition of these attendance data, the level of implementation for SIM-CERT professional development delivery may be higher than reported.

The professional development ratings were based on district-reported workshop attendance during both initial and ongoing training (refer to the exhibit below). Separate scores were assigned for initial and ongoing training, as planned, for each SIM-CERT teacher identified by the SR district implementation team. An adequate rating reflects attendance at *all* required professional development sessions while a not adequate rating indicates non- or partial attendance. The percentage of low and adequate ratings assigned to individual teachers on a case by case basis was calculated and disaggregated by district and cohort.

In Year 2, for scoring purposes, evaluators decided to use district records as the primary source of attendance data, which were made available in Year 2 by the SR district team. While the survey in Year 2 asked about attendance at professional development sessions, the primary purpose of the survey was to gauge teacher reported familiarity and use of the routines. It was determined that district records were a more reliable source of teacher attendance and therefore, this data source was used to assign actual scores for professional development. In the Year 1 report, descriptive information was provided but no scores were calculated.

As illustrated in Exhibit 60, of those teachers identified as participants in the training, a majority across districts and cohorts attended the requisite two days of initial training before the start of the academic school year. In addition, the percentage of teachers attending initial training sessions increased in Year 2 as compared to Year 1, in both districts.

Exhibit 60. Professional development ratings by district and cohort: Attendance according to model specifications

District/ Cohort	Initial Training (2 days total)		Ongoing Training Year 1 (2 days total)		Ongoing Training Year 2 (2 days total)	
	Adequate [a]	Not Adequate [b]	Adequate [a]	Not Adequate [b]	Adequate [a]	Not Adequate [b]
SPS [155] Cohort 1 (n=53)	46 (87%)	7 (13%)	1 [c] (2%)	52 (98%)	29 (74%)	10 (26%)
CPS Cohort 1 (n=48)	47 (98%)	1 (2%)	34 (71%)	14 [d] (29%)	37 (92%)	3 (8%)
SPS [156] Cohort 2 (n=80)	77 (96%)	3 (4%)			0	80 (100%)
CPS Cohort 2 (n=50)	50 (100%)	0			31 (62%)	19 (38%)

[a] Adequate = attendance at all required professional development sessions
[b] Not adequate = non or partial attendance at required professional development sessions
[c] This teacher received ongoing training during the 2006-2007 school year but was a Chicopee teacher during the first year of the grant.
[d] Two Xtreme Reading teachers did not attend ongoing training in Year 1.

Overall, initial attendance ratings were adequate across both districts. In Chicopee 98% of Cohort 1 teachers and 100% of Cohort 2 teachers received a rating of adequate for attendance at the planned initial training. In Springfield, the percentage of teachers who received adequate ratings for initial attendance was 87% in Cohort 1 and 96% in Cohort 2 respectively. Again, adequacy ratings are recorded only as presence at *all* initial training sessions. Teachers with partial attendance (i.e., teachers who attended one of the two days of initial training) were not counted as adequate.

[155] Three additional teachers appeared on the districts' professional development list for Year 1 but were never trained and five teachers on the list were not trained until after the school year was over effectively rendering them Cohort 2 teachers. Fourteen SIM-CERT teachers left the district by Year 2, one Xtreme Reading teacher left the district by Year 2. One of the remaining six Xtreme Reading teachers completed ongoing training in Year 2 and the other five did not. Nine SIM-CERT teachers left the district in Year 2. None were Xtreme Reading teachers.
[156] This number includes the five teachers from Year 1 who were not trained.

Relative to the ratings for attendance at the initial professional development trainings, those for developer-specified ongoing training were lower. In Springfield, 98% of Cohort 1 and 100% of Cohort 2 teachers did not receive ongoing training as planned during the academic year.[157] In Chicopee, 71% of Cohort 1 and 62% of Cohort 2 teachers attended all ongoing professional development sessions during their first year in the program and therefore received adequate ratings.[158] Finally, 74% of the Cohort 1 teachers in Springfield and 92% of those in Chicopee received an adequate rating for attendance at ongoing training during their second year of implementation.

The scores discussed above show variation between districts in the implementation of the professional development model. This variation is primarily a result of a delay in the provision of ongoing training (refer to Exhibit 61 which shows the schedule for professional development in Springfield-Chicopee for Year 1 and Year 2 of the whole school intervention). The Chicopee school district was able to use already scheduled in-service days to provide SIM-CERT training during the school year, however, Springfield was unable to do the same, which created differences between districts in training delivery schedules and had implications for scoring participation in professional development training. More specifically, although ongoing training as planned was to be provided during the school year (so that teachers could integrate what they had learned from professional development sessions in the classroom), Springfield provided ongoing training after the conclusion of the academic school year.

[157] Ongoing training received after the conclusion of the school was counted as ongoing training for the upcoming school year. More specifically, teachers who attended ongoing training in June/August 2007 were counted as receiving ongoing training for the 2007-2008 school year, not the 2006-2007 school year.

[158] During Years 1 and 2 of the grant, Chicopee provided one day of ongoing professional development in December and the other ongoing training day in March. On the in-service day in March the new cohort was trained in the Course Organizer, a routine that teachers implement for the following year's courses.

Exhibit 61. Sequence of SIM-CERT professional development activities by district

	Year 1 Implementation (2006-2007 school year)		Year 2 Implementation (2007-2008 school year)	
	Initial	Ongoing	Initial	Ongoing
Springfield	*Cohort 1* August 2006	*Cohort 1* June 2007		*Cohort 1* June/August 2008
			Cohort 2 June/August 2007	*Cohort 2* June/August 2008
Chicopee	*Cohort 1* August 2006	*Cohort 1* December 2006 March 2007		*Cohort 1* August 2007 March 2008
			Cohort 2 August 2007	*Cohort 2* December 2007 March 2008

During initial and ongoing SIM-CERT workshops over the course of 2 years, teachers were expected to receive content-based training in the following SIM-CERT routines: Unit Organizer, Course Organizer, LINCing, Framing, Concept Mastery, and Concept Comparison. According to developer- and district-provided documents, priority was given to the Unit Organizer routine as a foundational training for all to receive. Beyond what was considered foundational, teachers were also to be trained in Framing, LINCing, and Concept Mastery in Year 1. During Year 2, teachers were to receive training on Concept Comparison as well as other SIM-CERT components such as integrating units and links to literacy. Training in Concept Anchoring and other SIM-CERT routines was considered optional and left to teacher discretion.

According to district communications, SIM-CERT specifications regarding the professional development model lacked clarity and consistency across Years 1 and 2 of implementation.[159]

[159] The district team reported that developers were conscious of maintaining consistency of routine training across districts and across cohorts in what was planned and what was actually implemented.

The district team reported that developers stressed the importance of meeting the needs of the individual schools and districts, which led to fluctuations in the model as planned.[160] Specifically, developer plans for the routines SIM-CERT teachers would receive training in as well as the sequence in which these routines would be taught during Years 1 and 2 of the intervention did not remain constant over time. For example, there were conflicting messages regarding when teachers should receive training on the course organizer routine —ranging from ongoing training in Year 1, initial training in Year 1, or as part of ongoing training in Year 2. Similarly, contradictions are evident regarding the requirement that teachers receive training on the Concept Comparison and Concept Anchoring routines as well as training in Tool Box Development and Links to Literacy.

A combination of inconsistency in SIM professional development plans and district differences in professional development scheduling structures accounted for reported district variation in the types and amount of content (e.g., number of routines) teachers were exposed to during training sessions. District and developer documents indicate that Chicopee teachers received training in the use of the Course Organizer Routine in March of 2007,[161] while Springfield teachers did not receive this training until the conclusion of the 2006-2007 school year. Furthermore, Chicopee teachers initially received training in Framing, LINCing, and Concept Mastery while Springfield teachers received training in two of these three routines. According to documentation provided by the district's SR district implementation team, Springfield teachers received training in the Concept Mastery routine at the end of the school year. District differences in the sequence of professional development delivery and the content provided likely influenced survey responses and classroom model scores.

[160] It is difficult to distinguish between planned and actual as professional development plans were in often in development at the time of training.

[161] The Year 1 report stated that Chicopee trained teachers in the Course Organizer in August of 2006. Further review of district-provided documents indicated the Course Organizer routine was presented to Chicopee teachers in March of 2007.

Exhibit 62 presents Year 1 and Year 2 survey results for all teachers, across cohorts, reporting receipt of SIM-CERT training. Consistent with the professional development model as planned, survey data show that in both years, most SIM-CERT-trained teachers in Chicopee and Springfield received training in the Unit Organizer, LINCing, and Framing routines. According to Year 2 survey data, the fewest number of teachers reported receipt of training in Concept Comparison, which, according to district and developer-provided documents, was offered only to Cohort 1 teachers.

Exhibit 62. Number of respondents reporting participation in SIM-CERT routine by district

Survey Data source	District	Unit Organizer	Course Organizer	LINCing	Framing	Concept Mastery	Concept Comp.
Year 1	*CPS* (n=45)	43 (96%)	42 (93%)	31 (69%)	30 (67%)	12 (27%)	n/a
Year 1	SPS (n=45)	38 (84%)	18 (40%)	28 (62%)	22 (49%)	20 (44%)	n/a
Year 2	*CPS* (n=67)	53 (79%)	44 (66%)	42 (63%)	49 (73%)	40 (60%)	21 (31%)
Year 2	SPS (n=83)	67 (81%)	45 (54%)	53 (64%)	50 (60%)	39 (47%)	17 (21%)

In addition, more than half (67%) of survey respondents across districts who reported they were trained in SIM-CERT agreed that the SIM-provided professional development sessions prepared them to implement in the classroom.[162] The majority of SIM-CERT-trained survey respondents (67%) further indicated that they were generally pleased with the amount and quality of the professional development they have received.

[162] Categories of agree and strongly agree were collapsed for three items related to teacher satisfaction levels as reported above.

Focus group data and responses to the open-ended survey questions provided additional information regarding the perceived helpfulness of SIM-CERT training. Teachers tended to value content-specific professional development activities with trainers whose subject matter expertise matched their own. Cohort 1 teachers in Chicopee were generally critical of SIM trainings when trainers and training activities were disconnected from their content area.

Further, teachers across the two districts indicated that the hands-on exercises in the training that allowed collaboration with other teachers in their subject area were especially valuable to them. Teachers noted that they seldom have time to collaborate with other subject-area colleagues and that this training provided a rare opportunity to do so. One teacher explained:

> *So we picked up some good ideas and we were just happy that, yes…more science. And that we could all get together for once, because we don't always get to work with other science teachers in some of these workshops.*

A second component of the professional development specified by the model is school-embedded mentoring and support provided by a SIM-trained literacy coach. According to the five school-based literacy coaches interviewed, they provided a host of services to SIM-CERT-trained teachers in the school to supplement and build on what teachers were taught in the training sessions. Their primary role involved conducting informal classroom observations or visits to model, co-teach, and observe the implementation of SIM-CERT routines. Coaches typically debriefed with teachers one-on-one after their visits to discuss the level of fidelity in the implementation of instructional routines. They also provided feedback on lesson plans and the SIM-CERT devices teachers had created for classroom use, assisted teachers with using GIST software, and held monthly trainings after school. In addition to the implementation guidance provided by the school-based literacy coach, school administrators were also to provide support for SIM-CERT-trained teachers.

The requirement that literacy coaches possess Reading Specialist certification (or an "in process" designation) was eliminated given the lack of qualified staff available to fill these positions. In addition, districts reported that developers indicated that the literacy coaches did not need specialized training in reading instruction to be "good coaches." The districts reported working with the SIM-CERT team to identify appropriate teachers without a specialist certification to fill the literacy coach positions.

To ensure coaches and administrators were prepared to provide comprehensive support to SIM-CERT-trained teachers, the SIM-CERT model specified the provision of initial training for administrators and coaches. Additionally, onsite support and technical assistance were offered to coaches monthly on an as-needed basis.

Based on district records provided during Year 1, administrators and literacy coaches received initial SIM-CERT training as planned.[163] Administrators participated in a half-day initial training and orientation session in the fall of 2006. The five school-based literacy coaches participated in all developer-provided SIM-CERT training sessions in addition to the professional development activities indicated in the exhibit below.

Exhibit 63. District documentation: Participation in SIM-CERT training for literacy coaches

	DATE OF SESSION	DATE OF SESSION	DATE OF SESSION	DATE OF SESSION
	9/11-9/13	10/24-10/25	11/1-11/03	12/5-17/07
Number of literacy coaches participating in professional development (n = 5)	5	5	5	5
Number of hours	18 hours	12 hours	18 hours	18 hours
Basic features of instruction/activity	Instructional Coaching Institute	Completion of Instructional Coaching Institute	Book Study, Coaching Strategies, Routines	Book Study, Case Studies, Concept Mastery

[163] The districts did not provide documentation of additional training sessions for coaches and administrators in Year2 of the intervention.

According to Year 2 interview data, the coaches had mixed opinions on the usefulness of the ongoing mentoring support and technical assistance provided to them by the developers. Three coaches, across districts, felt supported by the SIM team. They cited the team's availability via email and phone, monthly visits, and helpful feedback and responses to questions. One coach stated that the SIM team representative did everything from observing them to having debriefings with them:

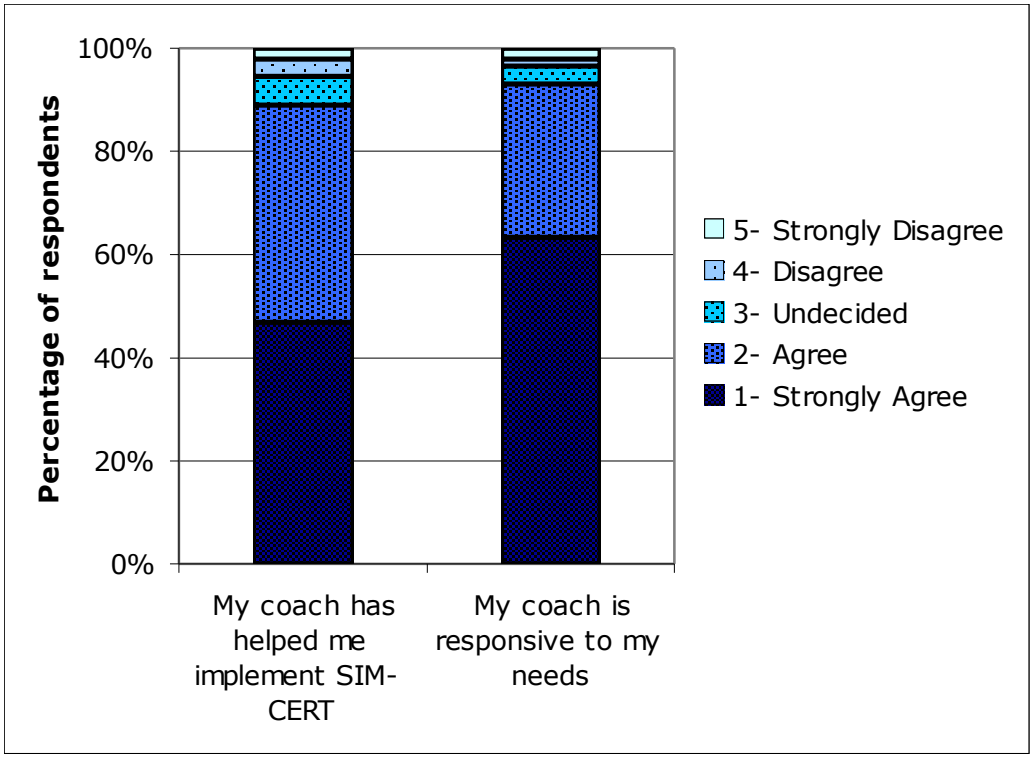
> *She has observed me do model lessons. She has gone with me on class visits… she's helped me reflect on a lot of issues surrounding coaching. She's provided me literature about coaching, books that I wanted to read.*

Two of the coaches, however, were less enthusiastic about the training and support they received from the SIM team, stating that while the team provided support frequently, the delayed response to email questions, vague explanations for future plans, and the didactic manner in which support was provided was frustrating.

On-site Mentoring Support

There was consensus among teachers and administrators that the support provided by the literacy coaches had been instrumental in the classroom-level implementation of SIM-CERT. Survey, focus group, and administrative interview data were overwhelmingly in agreement regarding the supportive role played by coaches in the implementation of SIM-CERT. Exhibit 64 presents teachers' opinions regarding the assistance coaches provided. Overall, survey data show high levels of agreement among teachers that the literacy coach helped them implement SIM-CERT routines (89%) and had been responsive to their questions and needs (93%).

Exhibit 64. Percentage of agreement among teachers that the school-based coach has been helpful and responsive (n=145)



Survey respondents and focus group participants were also in agreement that the level of implementation would have been lower without the coaches' continuous support and encouragement. One teacher stated:

> *The most important part of our professional development is having a coach right here with us to follow up on what we learned, helping us put it into practice. [Our school's literacy coach] has been an indispensable part of this process.*

Focus group participants further indicated that they might not have implemented SIM-CERT in the classroom without the presence and support of the literacy coach. Teachers noted:

> *I don't think SIM-CERT would have been supported if [our school's literacy coach] weren't here. It's one thing to work with young idealistic teachers who want to try anything. It's another thing to work with experienced teachers like us--we've seen programs come and go.*

> *She's always supportive and positive. In the beginning, I was hesitant. But she's always asking how she can help. You don't feel like you're bothering her at all. I feel very comfortable with her.*

Focus group participants attributed the coaches' positive influence on SIM-CERT implementation to their non-judgmental approach and their desire to help rather than evaluate SIM-CERT teachers. All five coaches confirmed that they used a non-evaluative approach to working with teachers. They emphasized that they attempted to build trust and rapport by being responsive to teachers' individual needs: co-teaching lessons, searching for classroom supplies, or even making copies and transparencies when teachers faced time restrictions. One of the coaches explained:

> *I do a lot of rapport building. I try really hard to be everyone's best friend. If a copier's jammed, I can help them out. If I'm walking by and I hear they need some pencils, I pick some up.*

The literacy coaches explained they attempted to *"lighten the burden of implementation"* for teachers by decreasing the number of planning steps involved in using the routines in the classroom.

As one literacy coach related:

> *I think the teachers have so many things to do, and there are so many initiatives, there's so much going around, that I think for some of them, if I can make it easier for them, maybe I'd find out what they're teaching and develop a little piece for one of the materials and what if we tried this, or what if we took this and worked on it together. So in some cases I tried to take a teacher with no involvement and brought them on board, and in some cases, I'd take somebody who was a little bit involved and tried to create a little bit more integrity with how they're using the materials. I'm always trying to support the people that are participating, and try to keep them involved.*

Interview data from school-level administrators mirrored the sentiments expressed by teachers (via focus groups and surveys). Administrators with various SIM-CERT implementation roles were positive about the support provided by the SIM-CERT literacy coaches. One administrator described the support the school's literacy coach has provided to SIM-CERT-trained teachers as follows:

> *The coach is fantastic. She's been excellent with my most challenging teachers and my best teachers. She's extremely supportive. She really has a way of connecting with them and supporting them through things. She's right there in the classroom with teachers all the time. If it weren't for her, I don't think it would be working.*

Two other administrators in charge of special education services in each district noted that the SIM-CERT literacy coach is *"very approachable"* and that the support provided has been *"wonderful."*

Classroom Level Implementation

At the end of the first year of implementation, minimum classroom model specifications were provided by developers requiring teachers to:

- Utilize a Unit Organizer for every unit taught for one course during the academic year
- Implement at least one additional routine for each unit (e.g., LINCing, Framing, Concept Mastery)
- Implement other routines as appropriate (refer to the SIM-CERT logic model presented in Section III)[164]

In addition to professional development, self-reported classroom level implementation was the second component of the SIM-CERT implementation ratings. Ratings were assigned based on survey responses regarding the use of SIM-CERT routines.[165] Ratings for the implementation of the classroom model are presented in Exhibit 65, disaggregated by district and cohort.

[164] According to district communications, SIM did not provide criteria for implementing the classroom model with fidelity until later in the first year of implementation. The expectations or criteria provided by SIM for the classroom model were minimal and much of classroom implementation was left to individual teacher discretion. The criteria used for scoring the implementation of the classroom model reflect this and include only the minimum developer-defined requirements.

[165] Scores for classroom usage of SIM-CERT routines were assigned according to teacher self-reports regarding the implementation of each routine at some point during the 2007-2008 school year. Scores did not take into consideration the frequency with which teachers implemented each routine in the classroom (i.e., whether teachers used a unit organizer for every unit taught) due to minimal information received from the developers on classroom model specifications during Years 1 and 2 of the intervention.

Exhibit 65. Classroom model ratings by district and cohort

Cohort	District	Met Minimum Usage Requirements Unit Organizer + 1 additional routine		Exceeded Minimum Usage Requirements Unit Organizer + 2 additional routines	
		Adequate	**Not Adequate**	**Adequate**	**Not Adequate**
1	SPS (n=20)	17 (85%)	3 (15%)	12 (71%)	5 (29%)
1	CPS (n=31)	31 (100%)	0	26 (84%)	5 (16%)
2	SPS (n=57)	38 (67%)	19 (33%)	24 (63%)	14 (37%)
2	CPS (n=33)	26 (79%)	7 (21%)	23 (88%)	3 (12%)
Total	(n=141)	112 (79%)	29 (21%)	85 (76%)	27 (24%)

Respondents who met the minimum developer-defined requirements as described above received a rating of adequate and those who did not received a rating of inadequate. Respondents who received a rating of adequate reported meeting minimum requirements; that is, use of the Unit Organizer routine plus one additional routine. Respondents who received a rating of inadequate for usage either used only the Unit Organizer routine or indicated that they had not used the Unit Organizer routine during the current school year. Ratings were not assigned to respondents with missing information regarding the Unit Organizer. A similar rating framework was also applied to determine which respondents exceeded developer-defined-routine-usage requirements. Thus, teachers who indicated they had used the Unit Organizer routine plus two or more additional routines received a rating of adequate and those who did not meet these criteria received a rating of inadequate.

According to the survey data, the majority of teachers across districts and cohorts met minimum requirements for classroom implementation of SIM-CERT. Higher percentages of Cohort 1 teachers reported adequate levels of implementation than Cohort 2 teachers.[166] Further, while classroom model scores were generally high across districts, a greater proportion of Chicopee teachers reported meeting minimum usage requirements than their Springfield counterparts in both cohorts. All Chicopee teachers in Cohort 1 reported meeting minimum usage requirements and 85% of Springfield teachers in Cohort 1 reported meeting minimum usage requirements. Similarly, a higher percentage of Cohort 2 Chicopee teachers reported meeting minimum usage requirements than Cohort 2 Springfield teachers.

Some evidence of district variation also emerged when examining the percentage of teachers exceeding developer requirements for classroom implementation. A greater proportion of Chicopee teachers in Cohort 1 (84%) and Cohort 2 (88%) reported exceeding usage requirements than Springfield teachers in Cohort 1 (71%) and Cohort 2 (63%). Higher levels of reported classroom-model implementation in Chicopee versus Springfield may be the result of district variation in the provision and schedule of professional development trainings and/or reported barriers to implementation (as discussed in the implication section). However, note that of the total number of SIM-CERT-trained survey respondents across districts and cohorts, 79% reported meeting minimum classroom model expectations.

Frequency of Implementation

Year 1 literacy coach interview results suggest that the frequency with which teachers implemented SIM-CERT was not uniform, particularly during at least the first four months of implementation. Some teachers reportedly used the routines regularly, while others were more hesitant to implement the intervention in their classrooms.

[166] According to district communications, classroom-level expectations for Chicopee teachers were more defined and were monitored by building administrators during Year 2. This may explain why fewer Cohort 2 Chicopee teachers report meeting classroom model requirements.

Examples of literacy coach feedback included the following:

> *Everyone except for one teacher is using the Unit Organizer. They are each using it to different levels. The group is pretty evenly divided with about a third writing the UOs [Unit Organizers] but not consistently implementing them, a third writing and implementing them with limited success, and a third writing and implementing them consistently with success—or at the highest level of fidelity.*
>
> *[Four are] doing it with a high level of fidelity. There are six more…planning and preparing to begin to use it with fidelity… I think it's better that they plan and do it well than do it without investment.*

Year 2 survey, focus group, and literacy coach data support these findings stated in the Year 1 report, in particular by revealing district variation in level of classroom model implementation. According to model specifications, SIM-CERT-trained teachers were expected to implement the Unit Organizer routine for every unit taught.

Exhibit 66 displays the reported number of units teachers planned using the Unit Organizer routine during the 2007-2008 school year.

Exhibit 66. Frequency of classroom implementation: Unit organizer

Reported frequency of use Unit Organizer	CPS	SPS	Total
	(n=59)	(n=56)	(n=115)
1-2 units	8 (14%)	25 (42%)	33 (29%)
3-4 units	19 (32%)	13 (22%)	32 (28%)
5 or more units	32 (54%)	18 (36%)	50 (43%)

Note: Percentages were based on the total number of teachers who report that they have used the Unit Organizer routine (i.e., valid percentage).

When asked the number of units planned using the Unit Organizer routine, the majority of Springfield teachers (42%) responded indicated "one to two units" whereas the majority of Chicopee teachers (54%) responded indicated "five or more units." Overall, the majority of respondents reported that they had implemented the Unit Organizer routine for five or more units throughout the 2007-08 school year.

Of those teachers using LINCing and Concept Mastery, a higher percentage of Chicopee teachers reported more frequent use than their Springfield counterparts (45% and 32% in Chicopee versus 34% and 28% in Springfield, respectively).[167] Teachers in Springfield reported more frequent use of the Framing routine within the past four weeks than Chicopee teachers (66% versus 45%, respectively).[168] Teachers reported similar usage of the Course Organizer routine across districts with the majority (75%) indicating they had used this routine to plan one to two courses during the 2007-2008 academic year.

Literacy coaches reported Chicopee teachers implemented SIM-CERT routines more frequently than their Springfield counterparts. While the Springfield coaches reported varying rates of implementation among SIM-CERT trained teachers, all literacy coaches noted that no more than one-third of their SIM-CERT teachers were frequently implementing the routines. The number of teachers who were not implementing the routines at all, as reported by Springfield's literacy coaches, varied widely by school from two-thirds to a small number of SIM-CERT trained teachers. In contrast, literacy coaches in Chicopee reported that nearly all trained teachers have used SIM-CERT. At both Chicopee schools, the literacy coaches reported that very few teachers, if any, were not implementing routines.

[167] Percentages were based on the number of teachers confirming use of LINCing and Concept Mastery routines (LINCing CPS = 31/67, SPS = 38/83; Concept Mastery CPS = 25/67, SPS = 18/83). Note numbers of teachers who responded to the item regarding frequency of use within the past four weeks (LINCing = 69; CM = 43) was similar to the number (LINCing = 71; CM = 45) who reported using CERT routines a minimum of one time during the academic school year.

[168] Percentages were based on the number of teachers who confirmed the use of the Framing routine in the Year 2 survey (CPS = 44/67; SPS = 38/83).

Quality of Implementation

In Springfield, literacy coaches indicated that approximately one-third of the SIM-CERT trained teachers were implementing the intervention with fidelity (i.e., implementing routines as intended). Two of the literacy coaches reported that more than one-third but less than half of SIM-CERT teachers was *"doing a pretty good job."* Another Springfield literacy coach was more critical and said that while one-third were implementing well, *"The rest are either not doing it or are pretending… one of two things. I don't think that this is being implemented to the degree that you could see an outcome."* In Chicopee, roughly two-thirds of the SIM-CERT- trained teachers were reportedly implementing the intervention well as per literacy coach feedback. One of the literacy coaches further claimed that, *"even my non-believers are doing it well."*

All of the coaches referred to using monitoring tools to document their interactions with teachers and to create a rough measure of each teacher's level of implementation. While three of the coaches referenced using a coaching scale developed by SIM staff (the Woodruff scale), the coaches collectively emphasized the use of the monitoring tools that they co-created themselves. These tools track the number of meetings with individual teachers; the frequency with which teachers submitted devices for feedback and review; the frequency of coach visits to teachers' classrooms to model the routines, co-teach, or observe SIM-CERT lessons; and the level of fidelity with which teachers created devices and implemented SIM-CERT routines in the classroom. Literacy coaches noted using these tools to gather data on SIM-CERT implementation for three basic purposes: (1) to measure levels of implementation for the SR district implementation team and/or SIM developers, (2) to reflect with individual teachers on how to more effectively apply SIM-CERT routines in the classroom, and (3) to determine the types of assistance that would be most relevant to an individual teacher's needs. Literacy coach interview data on frequency and quality of implementation are based on the measurements made from these monitoring instruments.

Focus group participants provided additional insight regarding when and why they choose to implement particular SIM-CERT routines.

Unit Organizer: Focus group participants in all five high schools indicated that they used the Unit Organizer routine for the following purposes: to prepare students for a test, to organize their own teaching and lesson planning, to give the big picture so students will know *"where they are going and where they have been,"* and to provide a structured format for student note-taking. Although nearly all teachers in the focus groups (across subject areas) indicated that they used the Unit Organizer routine, some were more enthusiastic about its usefulness than others. Several said that they had spent a substantial amount of time trying to understand the routine and found it challenging to determine how best to use it with students.

Framing: Teachers reported that they used the Framing routine in much the same way as they used the Unit Organizer routine that is, to help students take notes, study for tests, and understand a particular unit of study. The teachers who referenced the Framing routine generally agreed that this was one of the easiest routines to implement and that it required the least amount of preparation time. Math and science teachers referred to the Framing routine more often than other subject area-teachers, although several English teachers also said that they used Framing to help students structure their writing assignments. One teacher explained:

> *I use the Framing a lot. When I teach history and English, there's a lot of information my students are getting. The frame helps them organize the information and helps them organize their writing. It's the easiest routine for me, as a teacher, to use.*

LINCing: Responses from focus group attendants varied regarding the usefulness and practicality of implementing LINCs in their lessons. Science and math teachers generally indicated that they did not use LINCs in the classroom either because they believed it was not relevant to the subject matter they teach or because it took too much class time to implement. Comments from teachers included: *"LINC has no relevance in my subject area."* and *"I don't use LINCing. It takes too long."*

Another teacher added, *"LINCing takes days. I don't use it."* In contrast, several ELA and ESL teachers indicated that they used LINCing on a regular basis.

Other SIM-CERT routines: Very few teachers who participated in focus group sessions mentioned using the Course Organizer, Concept Mastery, or Concept Comparison routines. Those who did mention use of these routines did not provide specific information regarding when and how they used them with students. Focus group participants discussed the Unit Organizer, Framing, and LINCing routines most frequently in response to general questions regarding use.

Teacher-Reported Impacts on Student Outcomes

Given the phase-in plan for SIM-CERT, evaluators assume that should there be widespread impact, it will likely be gradual and as such the evaluation has been designed to track the progress of outcomes over time. Impact findings from the SIM-CERT portion of the Striving Readers initiative will be presented in future reports. However, survey analysis provides an early indication of future results related to the spread of SIM-CERT over time and to teacher perceptions of the impact of SIM-CERT on student achievement.

The exhibit below presents Year 1 and Year 2 data regarding the number of survey respondents who report receiving training in and using specific SIM-CERT routines.[169] This exhibit highlights the increased use of SIM-CERT routines over time. In accordance with developer requirements, the highest percentage of respondents reported using the Unit Organizer routine.

[169] Exhibit 64 depicts the total number and percentage of teachers who reported receiving training in as well as using CERT routines. Districts were combined to show change between Years 1 and 2 of implementation.

Exhibit 67. Number of survey respondents who report they have been trained and are using a routine [170]

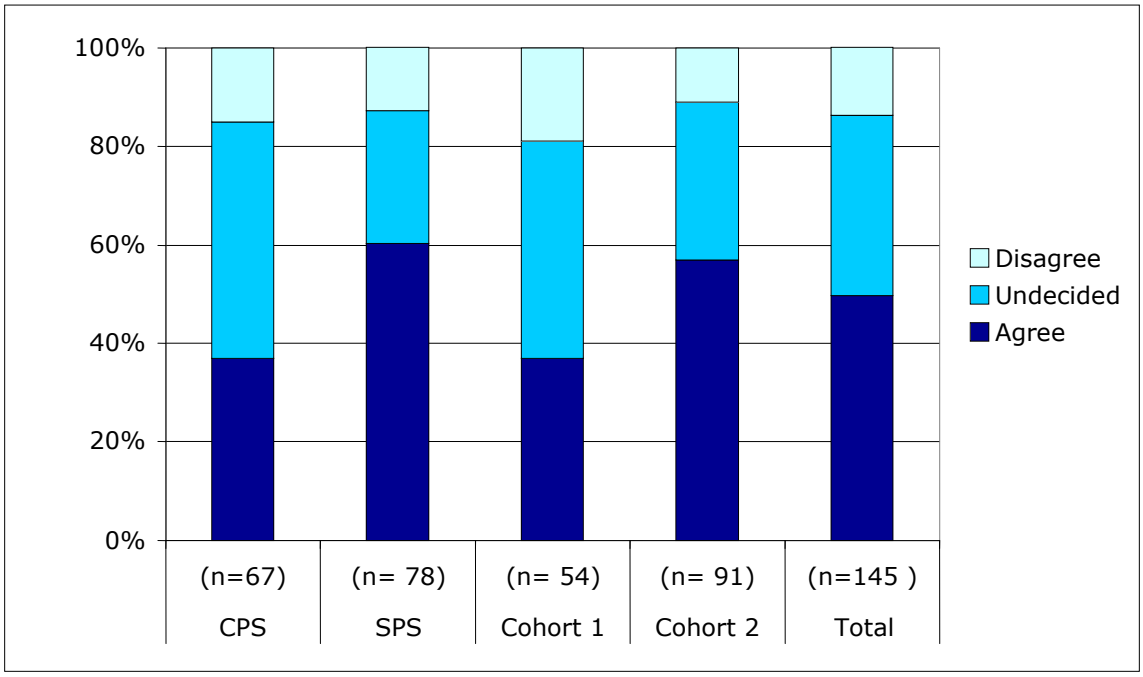
Data source	Unit Organizer	Course Organizer	LINCing	Framing	Concept Mastery	Concept Comparison
Year 1	78/81 (96%)	40/62 (65%)	37/59 (63%)	33/52 (63%)	13/32 (41%)	n/a
Year 2	106/120 (88%)	52/89 (58%)	61/95 (64%)	77/99 (78%)	41/79 (52%)	21/38 (55%)

Additionally, a majority of respondents stated that they have used the Course Organizer, LINCing, and Framing routines during both Year 1 and Year 2 of implementation. This exhibit shows increases over time in the number of teachers reporting that they are implementing SIM-CERT routines in the classroom. All of the percentages increase as well except for the Unit Organizer and Course Organizer routines. Absolute numbers are predicted to increase over time, especially for the planning routines such as Unit Organizer and Course Organizer. However, to show penetration over time, the percentage of the entire population of teachers should be investigated. Evaluators will look at this in the impact analyses that will be conducted in the future.

Exhibit 68 reflects teachers' perceptions regarding the contribution of SIM-CERT strategies to higher student performance levels. Positive perceptions may be a proxy for teacher buy-in. The greatest percentage of teachers either reported that they believed SIM-CERT strategies helped students understand course content or that they were undecided on this point.

[170] Denominator refers to the numbers reportedly trained in the routine.

Exhibit 68. Percentage of teachers who agree that SIM-CERT strategies help students better understand the course content (n= 145)



More specifically, Springfield teachers (60%) and Cohort 2 teachers (57%) across districts primarily agreed that SIM-CERT strategies helped students gain a better understanding of course content; nearly half of Chicopee teachers (48%) and Cohort 1 teachers (44%) were undecided. Less than 15% of SIM-CERT-trained teachers across districts and cohorts disagreed or strongly disagreed with the contention that SIM-CERT strategies facilitated student learning. Springfield's reliance on voluntary rather than mandatory recruitment efforts for participation in Cohort 2, a variation from the model as planned, could potentially be associated with higher levels of agreement among Springfield versus Chicopee (and Cohort 1) teachers that CERT routines have a positive impact on student learning outcomes. That is, if teachers were more motivated to participate, they may be more motivated to perceive student improvement and, therefore, to report positive impacts.

However, the inclusion of voluntary teachers (i.e., more motivated to participate) presents challenges in interpreting any overall study outcomes.[171]

In general, responses to focus group questions as well as open-ended survey questions mirrored responses to survey questions regarding perceptions of the efficacy of SIM-CERT strategies for improving student understanding (i.e., many teachers either offered positive feedback or indicated that they were undecided). A smaller number of respondents reported they were skeptical about the effectiveness of SIM-CERT strategies. Examples of focus group and open-ended survey feedback included the following:

> *I think it is very good at helping to organize units I teach. In turn, this helps my students understand content better.*
>
> *I am still undecided. The jury is still out. My opinion is still being formulated.*
>
> *I use it. I do. But I don't know if I am seeing results that are positive. This is one of the lowest grades I have ever had. I don't see it reflected in the grades.*

Implications

The following section summarizes differences between the proposed and planned intervention and describes facilitators as well as barriers experienced in the implementation of SIM-CERT.

Professional Development Model

Delayed provision of ongoing training sessions in Springfield. The Springfield and Chicopee districts initially proposed conducting similar professional development trainings together and at the same time intervals.

[171] As described previously, the recruitment was to be systematic given the majority of teachers were to be trained before the grant ended. Mandatory recruitment with selection in groups identified represented both an equitable approach as well as prevented threats of selection bias (for any observed outcomes).

However, due to differences in the professional development scheduling at the district level, Chicopee provided ongoing training as planned during the school year and Springfield offered ongoing professional development after the school year in which the routines taught in the professional development trainings were to be implemented. *The lower ratings for the professional development model in Springfield versus Chicopee are a direct result of this lag in the provision of ongoing training.* Specifically, while Chicopee offered ongoing professional development as planned roughly three and six months following the initial professional development sessions, Springfield offered follow-up training nine months after initial training and after the close of the school year.[172] This was the case for both Year 1 and Year 2 of implementation and constitutes a significant difference in planned fidelity to the model between Springfield and Chicopee districts.

Additionally, due to the delayed provision of ongoing professional development in Springfield, Cohort 1 teachers did not receive training in the Course Organizer and Concept Mastery routines until the conclusion of the 2006-07 school. Chicopee teachers in both cohorts received training in the routines as planned according to district documentation. Differences between the districts in the timing of offered professional development and the number of routines offered will likely have an impact on reported levels of implementation in the classroom; that is, if teachers have not been trained in a particular routine, they were not likely to have reported that they implemented that routine in the classroom.[173]

Teacher attrition. Teacher attrition may also constitute a barrier to SIM-CERT implementation, especially in Springfield. Districts' first and second year training rosters yielded information regarding the attrition of SIM-CERT teachers from the first to the second year of the program. In Chicopee for example, of the forty-eight teachers that received SIM-CERT training in the first year, nine were no longer at the district in Year 2.

[172] In some cases, teachers received ongoing training for the 2006-07 and 2007-08 school years in August of 2007 and 2008 respectively.

[173] As already noted, the Course Organizer routine is presented to Chicopee teachers during ongoing training in March, but not required to be implemented until the following fall. Thus, one of the two days of ongoing training in Chicopee would likely have little impact on classroom use of SIM-CERT routines during the school year in which training was received.

In Springfield, of the fifty-two teachers that received SIM-CERT training in the first year, fourteen were no longer working in the district and two did not attend training in the second year despite the fact they were still teaching at a participating high school in the district. In at least three cases, the district acknowledged that Cohort 1 teachers had attended training in Year 2 though they were absent from the district's Year 2 training roster.

As a school-level administrator commented:

> *I think one of the things I might be curious about is, I know at the end of a year, some of our teachers leave our building, so I'm assuming other teachers will as well. If there (are) 25 teachers at the beginning of the year, and that's at the beginning of the year, and there's a cohort, it has been reduced at the end. That may affect how [implementation] looks from the district perspective.*

Literacy coaches are a key support for SIM-CERT implementation. The Springfield-Chicopee Striving Readers implementation plan requirement that qualified literacy coaches should have a Reading Specialist certification or an "in process" designation was not realized for two reasons: (1) the lack of available and qualified staff and (2) because developers indicated to districts that Reading Specialist certification was not required. Districts worked with the SIM-CERT team to identify and train teachers without this specialist certification.

Despite being technically off-model in that the literacy coaches did not all possess reading specialist certification, multiple data sources indicate the school-embedded literacy coaches constituted a key support for the implementation of SIM-CERT. Teachers and administrators alike stated that the intervention would not be implemented to the current degree of fidelity without the support of the literacy coaches. This is not surprising, considering the literature on effective professional development models (Guskey, 2000).

Classroom Model

Expectations of the developer. Springfield and Chicopee reported that they had not received specific information regarding developer expectations for SIM-CERT implementation during Year 1.

Minimum teacher requirements for implementation were provided by SIM later in the first year. These requirements included developing a Unit Organizer for every unit delivered, and the implementation of a routine with each unit and other units as appropriate.[174] Specifications beyond this were not provided as decisions regarding how and which routines should be implemented were left to teacher discretion.[175]

Classroom-level implementation for Year 2. According to self-report data from the Year 2 survey, more than half of the teachers across districts and cohorts met the minimum requirements for classroom implementation of SIM-CERT. Higher percentages of Cohort 1 teachers reported adequate levels of implementation than Cohort 2 teachers. A greater proportion of Chicopee teachers claimed to meet and exceed minimum usage requirements than their Springfield counterparts in both cohorts. Springfield generally had lower levels of classroom implementation according to multiple data sources. Year 1 interview findings suggest that regular use of SIM-CERT routines varied according to the type of routine as well as levels of teacher buy-in and reported training, a finding reiterated by Year 2 data.

The following interrelated barriers to implementation of the whole school intervention were reported: (1) Selection and recruitment processes for inclusion in SIM-CERT cohorts; (2) structural issues regarding time for planning, instruction, and collaborating with CERT-trained peers; and (3) lack of administrative knowledge and support for SIM-CERT. Although these barriers were cited by teachers, coaches, and administrators across districts, they were more evident in Springfield and therefore may have contributed lower ratings in professional development and classroom-level implementation in that district.

[174] CLC Program Evaluation Implementation Phase Tool Kit, February 2007.
[175] No benchmarks have been provided to evaluators regarding the expected percentage of teachers who should be using specific routines regularly during the first and second years of implementation.

Selection and recruitment processes. In Year 1 and Year 2 interviews, coaches indicated that a lack of buy-in by teachers to the SIM-CERT program contributed to varying levels of classroom-level implementation. Coaches explained that many teachers had not volunteered to participate in the intervention and were thus less willing to use the routines in their classrooms. As one coach stated, *"They haven't refused. They just haven't bought in."* Year 2 focus group data provided teacher perspectives on the link between selection criteria and methods for inclusion in SIM-CERT cohorts and their level of buy-in to the program. Focus group members from both districts indicated that teachers were removed from the trainings they had planned to attend in the first year of the intervention and were told, with "short notice," that they would be attending SIM-CERT training instead. Focus group participants stated that this approach led to resentment (i.e., low levels of teacher buy-in) among Cohort 1 teachers.

Cohort 2 teachers in Springfield reported that they were given the choice to participate in SIM-CERT training and explained that this change helped to increase their level of buy-in to the intervention. However, teachers in both districts were to be trained over the course of the grant's implementation, as proposed. Perceptions that SIM-CERT training is optional for teachers over the five years of the grant may illustrate the extent to which teachers and school staff were not aware of the district's commitment to train *all* teachers as a requirement of a *school-wide* content literacy model. As per the original district plan, teachers were to be selected according to pre-specified criteria, not on a volunteer basis for reasons previously stated. While the teacher-selection strategy reportedly changed to a volunteer-based selection process in the second year of the intervention in Springfield, this was not the case in Chicopee. Chicopee implemented SIM-CERT training as planned using the designated selection criteria and the planned training schedule. As participation in SIM-CERT was voluntary in Springfield and training was not a component of the in-service professional development, incentives were used in the spring to recruit teachers.

Time for planning, instruction, and collaboration. As indicated by survey, focus group, and literacy coach responses, several structural issues were perceived as barriers to the implementation of the whole school intervention: (1) limited amount of instructional time in which to implement SIM-CERT routines, (2) limited amount of time in which to prepare lessons that integrate SIM-CERT routines, and (3) few opportunities to collaborate with other SIM-CERT-trained teachers.

Springfield teachers and literacy coaches expressed these concerns more than teachers and coaches in Chicopee, and it was suggested that these issues could be remedied by increased administrative knowledge of and support for the intervention.

Teachers participating in focus groups across the five high schools identified time limitations as the most serious issue inhibiting classroom implementation. Focus group participants expressed their concern that SIM-CERT implementation *"takes an awful amount of time"* and that they were doing *"a lot of work"* in preparing SIM-CERT devices, developing SIM-CERT-related materials, and integrating routines into lesson plans. Multiple teachers across both districts and cohorts who participated in focus groups and/or responded to the survey said they did not have the necessary time to prepare SIM-CERT routines for their classes. For example, one survey respondent said, *"More time is needed to incorporate these strategies into the everyday classroom. There is not enough time to adequately prepare these strategies for use in all units in all classes."* In addition to the amount of planning time required to incorporate the SIM-CERT routines, focus group participants and survey respondents indicated that they lacked a sufficient amount of time to integrate SIM-CERT into already limited instructional time with students. In other words, teachers saw curricular demands as competing with the time needed to implement SIM-CERT routines in the classroom throughout the academic year. As one teacher explained, *"Well, the amount of time needed to cover content conflicts with the time needed to implement all the SIM-CERT materials. It's a struggle."*

Another teacher concurred, *"I love [the routines] but I don't have the time to incorporate them into my classes the way I want."* Across both districts, 38% (n=55) of survey respondents agreed that the time needed to implement SIM-CERT routines conflicted with the time needed to cover course content throughout the year. Further, nearly half or 49% (n=71) of survey respondents indicated that they lacked sufficient planning time to incorporate SIM-CERT routines into their classes.

Teacher collaboration is a critical component of the implementation of a school-wide initiative. The literature on school improvement illustrates that the provision of opportunities for teachers to share what they have learned can have an empowering effect on staff and can create more belief in and support for the changes being implemented. In a sense, schools can begin creating their own professional learning communities.[176]

The majority of teachers indicated via survey and focus groups that collaboration with other SIM-CERT-trained teachers was valued but opportunities for such collaboration were not always available. A lack of co-planning and collaboration time was viewed by some to be the result of institutional constraints, which reportedly affected Springfield teachers more than their Chicopee counterparts. Springfield teachers who participated in the focus groups indicated that they had few opportunities to collaborate with other SIM-CERT-trained teachers beyond the ongoing trainings facilitated by SIM staff. They further stated that the barriers to collaborating with their peers stem from administrative and school-level structural issues: "*It's hard to collaborate because the school is financially strapped and it's hard to get coverage. At meetings we don't collaborate. We get lectured and told what to do."* Teachers at another Springfield school felt there was not enough time for *"teachers to collaborate on a regular basis.*"

[176] Sergiovanni purports (2000) "developing a community of practice may be the single most important way to improve a school" (p.139).

Chicopee teachers reported more collaboration time and indicated it was built into their schedules. In Chicopee a "community piece" was included in the implementation expectation checklist, according to one literacy coach, to require collaboration as a part of the plan. Teachers were encouraged and expected to invite a teacher into their classroom or to visit another SIM-CERT classroom to observe. At one of the schools in Chicopee, focus group participants stated that they shared lesson plans and devices electronically: *"Any lesson we've done is available to all of us (i.e., SIM-CERT teachers). I end up taking someone else's but tweaking it for myself."* In addition, they explained they were compiling a SIM-CERT library as a resource for other SIM-CERT teachers.

Administrative knowledge and support of SIM-CERT. According to literacy coach interviews and focus group data, district variations in the amount of administrative knowledge and support of SIM-CERT may have impacted the level of implementation. In Springfield, data indicate there is a perception of low levels of administrative support. School staff reported that administrators at the school- and district- level did not have specific knowledge about SIM-CERT. In particular, it was reported that administrators did not consider SIM-CERT a priority and that there was not enough support and follow-up for this intervention. For example, Springfield focus group participants explained that their performance evaluations, conducted by administrators, did not include SIM-CERT teaching practices and were not aligned with expectations for SIM-CERT classroom-level implementation. School staff also mentioned that basic expectations for SIM-CERT classroom-level implementation needed to be more clearly articulated. None of the teachers or the literacy coaches mentioned that methods for accountability to check on follow-through of CERT implementation were being used.

The lack of administrative knowledge and support of the whole school intervention in Springfield may be attributed to high levels of administrative turnover during Years 1 and 2 of implementation. Of the six administrative positions at the school level (principal and assistant principal), Springfield saw eleven staff members fill these positions in a total of two years.

For example, one Springfield high school had three principals in two years. In contrast, the two high schools in Chicopee retained the same principal and assistant principal for Years 1 and 2 of the grant. SIM-CERT training and orientation for school-level administrators was held only once in the fall of 2006 prior to Year 1 implementation. Due to high rates of administrator turnover in Springfield, many of the current administrators have likely not received SIM-CERT training and thus, may lack knowledge of how the whole school intervention is to be implemented.

Low levels of administrative knowledge and support for the program as reported by focus group participants and literacy coaches may have also precipitated the manner in which teachers were selected for and informed about their participation in SIM-CERT.[177] According to district communications, Springfield teachers were recruited in August of 2007 with short notice for inclusion in Cohort 1. Teachers were recruited for inclusion in Cohort 2 on a volunteer basis, despite initial plans to include SIM-CERT training in required district professional development. There is evidence that the SR district implementation team made repeated outreach efforts to provide information and explain what was and would be occurring, but less is known about efforts to follow-up at the school level. In addition, contractual and union concerns often arose throughout the first two years of the grant.

In Chicopee, data sources indicate there were moderate to high levels of administrative support for the whole school intervention and that, to some extent, school-level administrators have communicated the importance of SIM-CERT. At one high school in Chicopee, one school staff member indicated that *"strong administrative support"* exists and that administrators communicate at faculty meetings and professional development days the importance of content enhancement routines *"what's expected [of teachers] and what's coming up usually in a positive way."* It was reported that attendance at professional development sessions by administrators *"sends a message"* about the importance of the initiative.

[177] Other factors such as the short time frame between grant award and implementation as well as conflicts with the Control Board in Springfield may have contributed to district variation in teacher recruitment for SIM-CERT training in Year 1.

VIII. Evaluation Summary

The Springfield and Chicopee school districts have overcome many obstacles in the development, planning, and implementation of their Striving Readers (SR) grant. In particular, two dissimilar districts have implemented two targeted interventions (all other SR grantees implemented only one) as well as one whole school intervention.

Many of the barriers presented in the implementation of the grant in the first year resulted from both contextual and contractual factors, which did not necessarily emerge from the intervention models themselves but may have resulted from attempts to fit the models as required into this context (refer to the logic models for an overview of context). Some of the contextual factors included: the urban setting, population, and student needs; the various policies of the schools and districts addressing scheduling, administrative issues, etc.; as well as general staffing and personnel matters.[178] Contractual complexities specifically refer to the requirements for the grant implementation; the monitoring and oversight of the fidelity of implementation; and the observance of the rigorous research specifications.

Given the challenges inherent in both creating a successful collaboration between two districts and implementing two interventions, it is not surprising that complexities arose which would not normally be encountered in a standard literacy program implementation. An initial barrier related to the rigorous research requirements, for example, involved the cooperation, ability, and willingness of both districts to incorporate a "true" control group to address the counterfactual (i.e., *what would happen in the absence of treatment*). Additional challenges involved the need to standardize implementation across two very different district and school systems.

[178] One of the districts SR program leads took another position elsewhere prior to the first school year of grant implementation.

Intervention plans necessitated consistent tailoring to accommodate rigorous research study requirements and unanticipated time by district staff and evaluators was spent to ensure successful implementation. At the same time, districts faced changes in lead program staff, challenges related to communication with stakeholders and participants, and complications in screening and placing the population of students who were randomly assigned to participate in the targeted interventions.

Progress was made in overcoming these barriers, particularly in Year 2. Districts implemented each of the targeted interventions while maintaining the integrity of the randomized controlled trial design and assignment and repeatedly demonstrated their commitment to ensuring the success of the grant. District staff collaborated fully with evaluators in all phases of the evaluation. Their serious consideration of any potential positive or negative influences on study outcomes as well as "full disclosure" has been commendable. Such diligence ensures that the final results of this study will produce information that can be used by policymakers, district administrators, and school staff to make confident choices regarding effective interventions for their students.

References

Bloom, H.S., Richburg-Hayes, L., & Rebeck Black, A. (November, 2005). *Using covariates to improve precision: Empirical guidance for studies that randomize schools to measure the impacts of educational interventions.* MDRC Working Papers on Research Methodology. Manpower Demonstration Research Corporation New York, N.Y. www.mdrc.org last retrieved December 1, 2005.

Bloom, H.S. (2004, March). *Randomizing groups to evaluate place-based programs.* MDRC Working Papers on Research Methodology. Manpower Demonstration Research Corporation New York, N.Y. www.mdrc.org last retrieved December 1, 2005.

Content Learning Center, Kansas University (February, 2007). CLC Program Evaluation Implementation Phase Tool Kit. Author.

Corrin, W., Somers, M.-A., Kemple, J., Nelson, E., & Sepanik, S. (2008). *The Enhanced Reading Opportunities study: Findings from the second year of implementation* (NCEE 2009-4036). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Faddis, B. (personal communications, November, 2007). Evaluation materials developed based on information provided by Dr. Suzanne Robinson, University of Kansas, Center for Research on Learning, November, 2007). RMC Research, Portland, Oregon.

Fixsen, D.L., S.F., Blasé, K. A., Friedman, R. M. & Wallace, F. (2005). *Implementation research: A synthesis of the literature.* Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Goodson, B. (2006, October). *Measuring implementation.* Presentation at the U.S. Department of Education Striving Readers Program Local Evaluators Conference, Washington, D.C. Abt Associates, Inc.

Guskey, T. (2000). *Evaluating professional development.* Thousand Oaks, CA: Corwin Press.

Hasselbring, T. S., & Goin, L. I. (2004). Literacy instruction for older struggling readers: What is the role of Technology? *Reading and Writing Quarterly*, 20(2), 123-144.

Learning Point Associates (2008). *CSR Practitioner's Guide to SBR - What is scientifically based research?* The Center for Comprehensive School Reform, Learning Point Associates. Retrieved January, 2008. http://www.centerforcsri.org/pubs/pg/sbr.htm.

Massachusetts Department of Education. (October, 1999). *Massachusetts Comprehensive Assessment System.* 1998 MCAS Technical Report Summary. http://www.doe.mass.edu/mcas/1998/techrpt_sum.pdf

Meltzer, J. & Bray, C. (2006, October). *Core components and instructional practices of reading support in classrooms for middle and high school struggling readers: Summary chart for instrument development/design.* Presentation at the U.S. Department of Education Striving Readers Program Local Evaluators Conference, Washington, D.C. Center for Resource Management-Public Consulting Group, Inc. (CRM-PRG).

Meltzer, J. (2006, October). *Measuring treatment-control differences in literacy instruction: Issues, considerations and ideas.* Presentation at the U.S. Department of Education Striving Readers Program Local Evaluators Conference, Washington, D.C. Center for Resource Management-Public Consulting Group, Inc. (CRM-PCG).

Murphy, R., Penuel, W., Means, B., Korbak, C., & Whaley, A., & Allen, J. (2001). *E-desk: A review of recent evidence on the effectiveness of discrete educational software*. Menlo Park, CA: SRI International.

No Child Left Behind Act of 2001, Pub. L. no. 107–110 (2002). http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf.

Orr, L. (1999). *Social experiments: Evaluating public programs with experimental methods.* Thousand Oaks, CA: Sage.

Raudenbush, S.W., Martinez, A., & Spybrook, J. (November, 2005). *Strategies for improving precision in group-randomized experiments. W.T. Grant Foundation's Building Capacity for Evaluating Group-Level Interventions.* www.wtgrant.org Retrieved December 1, 2005.

Raudenbush, S. W., Spybrook, J., Liu, X., & Congdon, R. (2004). *Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software.* Last retrieved January, 2006 from http://sitemaker.umich.edu/group-based/files/od_documenattion/chapters1-8_vmarch09-05.pdf.

Redfield, D. (January, 2004). *An evaluator's guide to scientifically based research.* Institute for the Advancement of Research in Education at the Appalachia Educational Laboratory (AEL). Retrieved December, 2007. http://www.thefreelibrary.com/An+educator's+guide+to+scientifically+based+research:+documenting...-a0112313352.

Schmidt, S., & Kingman, M. (2006, November 15). Message posted to Striving Readers electronic mailing list.

Schochet, P. Z. (2005). *Statistical power for random assignment evaluations of education programs*. Princeton, NJ: Mathematica Policy Research, Inc.

Scholastic Reading Inventory. New York, NY: Scholastic, Inc.

Scholastic, Inc. (n.d.) Instructional model. Retrieved October 20, 2007, from Scholastic.com: http://teacher.scholastic.com/products/READ180/overview/instrmodel.htm.

Scholastic Professional Paper (March, 2006). Because you can't wait until spring: Using the SRI to improve reading performance. Kimberly A. Knutson, School District of Palm Beach County. New York, NY: Scholastic, Inc.

Scholastic, Inc. (2005a). READ 180 teacher implementation guide. New York, NY: Scholastic, Inc.

Scholastic, Inc. (2005b). READ 180 leadership implementation guide: Supporting READ 180 in your district. New York, NY: Scholastic, Inc.

Scholastic, Inc. (2005c). *READ 180 America's premier reading intervention program for elementary through high school.* New York, NY: Scholastic, Inc.

Scholastic, Inc. (2005d). *READ 180 participant's guide: Using the SRI and the Lexile Framework effectively with READ 180.* New York, NY: Scholastic, Inc.

Scholastic, Inc. (2005e). READ 180 Placement, Assessment, and Reporting Guide. New York, NY: Scholastic, Inc.

Scholastic, Inc. (2004). *READ 180 research protocols and tools*. New York, NY: Scholastic, Inc.

Scholastic, Inc. (2002). Scholastic's READ 180: A heritage of research. Davidson, J., & Miller, J. (Scholastic Eds.). New York, NY: Scholastic, Inc.

Sergiovanni. T. (2000). *The lifeworld of leadership: Creating culture, community and personal meaning in our schools.* San Francisco, CA: Jossey-Bass.

Stanford Diagnostic Reading Test, 4th Edition (SDRT-4) Technical Manual. Original SDRT4 Author: Bjorn Karlsen and Eric Gardner, 1995. Harcourt Assessments, Inc.

UMASS Donahue Institute (August 2008). *An evaluation of the Commonwealth Pilot School Initiative. Year one report: Intermediate outcomes of the Commonwealth Pilot School*

Model. Retrieved November 1, 2008 from: www.doe.mass.edu/research/reports/eval.html

U.S. Department of Education. (2007a). Striving Readers: Purpose. Washington, DC: U.S. Department of Education. Retrieved from: http://www.ed.gov/programs/strivingreaders/index.html.

U.S. Department of Education. (2007b). Striving Readers: Funding status. Washington, DC: U.S. Department of Education. Retrieved from: http://www.ed.gov/programs/strivingreaders/funding.html

Waxman, H. C., Connel, M. L., & Gray, J. (December, 2002). *A quantitative synthesis of recent research on the effects of teaching and learning with technology on student outcomes*. Naperville, IL: North Central Regional Education Laboratory. Retrieved January 5, 2006, from http://www.ncrel.org/tech/effects/index.htm